Omicron, Delta Waves and Beyond: Variant-specific peptide panels for diagnosis, characterization and surveillance of emerging SARS-CoV-2 variants during the pandemic

Subina Mehta[1], Valdemir M. Carvalho[2], Andrew T. Rajczewski[1], Björn Grüning[3], James E. Johnson[1], Reid Wagner[1], Olivier Pible[4], Jean Armengaud[4], Timothy J. Griffin[1],and Pratik D. Jagtap[1]

1. University of Minnesota, Minneapolis, MN; 2. Division of Research and Development, Fleury Group, São Paulo, SP; 3. University of Freiburg, Georges-Köhler-Allee Brazil Department of Computer Science, Freiburg; 4. CEA Life Sciences Division, DSV, IBEB, SBTN, Bagnols-sur-Cèze, France.

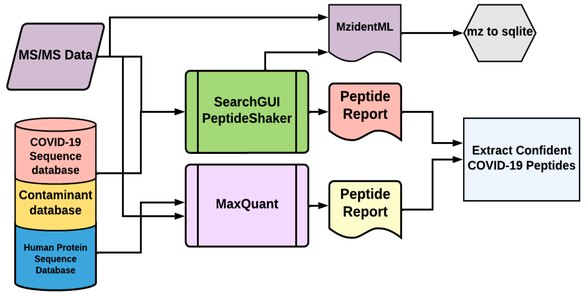
INTRODUCTION

- The COVID-19 pandemic continues with the emergence of new strains.
- As virulence seems to be increasing with each variant, there is a dire need to advance the MS-based diagnostics to identify optimal strain-specific biomarker peptides from non-invasively collected clinical samples.
- The Galaxy-P Team has developed workflows to detect and validate the SARS-CoV-2 peptides useful for developing targeted MS-based biomarker assays.

EXPERIMENTAL METHODS

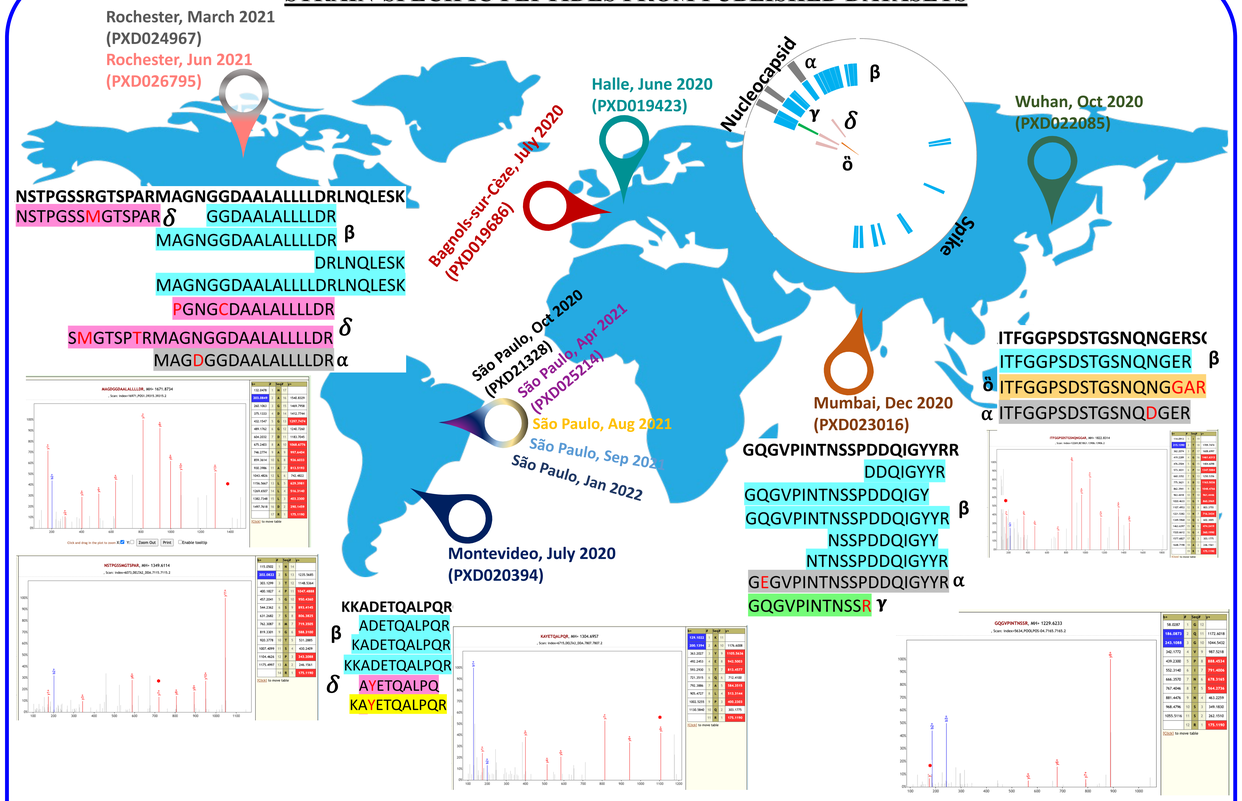
- Publicly available MS datasets from patient samples were searched against SARS CoV-2 proteome (including the variant sequences) using a DISCOVERY WORKFLOW within Galaxy Platform (see below).
- The DISCOVERY WORKFLOW detected SARS CoV-2 peptides from the clinical samples.
- Our VALIDATION WORKFLOW was used to verify the peptide detection and strain-specificity in the clinical samples.

DISCOVERY WORKFLOW



- MS/MS data from publicly available & nasopharyngeal patient samples were searched against protein FASTA files from SARS-CoV-2 Wildtype (WT) & variant strains along with Human UniProt & contaminants using SearchGUI/Peptide Shaker & MaxQuant search algorithms.
- Confident peptides were extracted from the peptide results & SARS-CoV-2 peptides were filtered by removing human proteins & contaminants.
- We detected 105 SARS-CoV-2 peptides from twelve clinical datasets.

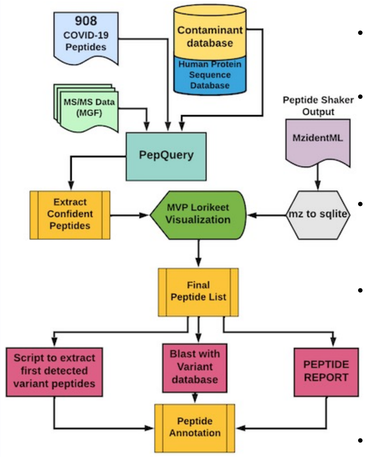
STRAIN-SPECIFIC PEPTIDES FROM PUBLISHED DATASETS



Peptides	Germany July 2020	France July 2020	Uruguay July 2020	Brazil Oct 2020	China Oct 2020	India Dec 2020	US Mar 2021	Brazil Apr 2021	US Jun2021	Brazil Aug 2021	Brazil Sep 2021	Brazil Jan 2021
MAGDGGDAALALLLLDRLNQLESK												
DGIIWVATEGALNTPK												
GQGVPINTNSSR												
NSTPGSSMGTSPAR												
SMGTSPTRMAGNGGDAALALLLLDR												
ITFGGPSDSTGSNQNGGAR												

α Alpha (B.1.1.7) β Beta (B.1.351) B.1.36 γ Gamma (P.1) δ Delta (B.1.671.2) ö Omicron (BA.1)

VALIDATION WORKFLOW



- 105 peptides were added to the 803 peptides detected from previous studies, including from our study[1].
- The 908 peptides were verified for their detection in MS datasets using PepQuery Tool.
- The PepQuery Tool ranks the matches of the peptide panel to COVID-19 MS data against a reference proteome containing proteins from human and contaminants.
- Resulting data was filtered to generate a list of confident peptide matches with associated P-values to SARS-CoV-2 and its variants.
- The peptide list was further analyzed using bioinformatic tools such as BLAST-P and Unipept. The spectral quality of these verified peptides was inspected using Lorikeet tool within the Multi-omics Viewing Platform (MVP)[2].
- Verified peptide were further investigated to extract additional information using the GSAID repository

CONCLUSION

- Most of the verified strain-specific peptides, align to the nucleocapsid, and spike proteins of the viral proteome and would serve as the optimal targets for direct detection of SARS-CoV-2 strains.
- The flexibility of the Galaxy workflows has the potential to detect the emerging strains on newly published MS datasets (DDA or parallel-reaction-monitoring (PRM)). We are expanding our peptide panel to include peptide targets to detect emerging strains.
- The Galaxy workflows along with the input data and updated results are available via https://covid19.galaxyproject.org/proteomics/.

ACKNOWLEDGEMENTS

Funding: National Cancer Institute - Informatics Technology for Cancer Research (NCI-ITCR grant 1U24CA199347) and National Science Foundation (grant 1458524).
Computational Resources: Minnesota Supercomputing Institute and The Galaxy Europe.

REFERENCES

1. Rajczewski, A.T., Mehta, S., Nguyen, D.D.A. *et al.* A rigorous evaluation of optimal peptide targets for MS-based clinical diagnostics of Coronavirus Disease 2019 (COVID-19). *Clin Proteom* 18, 15 (2021). https://doi.org/10.1186/s12014-021-09321-1
2. McGowan. T, Johnson. J et al. Multi-omics Visualization Platform: An extensible Galaxy plug-in for multi-omics data visualization and exploration. Gigascience (2020) https://doi.org/10.1093/gigascience/giaa025