Representative SARS-CoV-2 genomic sampling strategies allow accurate inference of **lineage-specific symptom prevalence** and delivers measures of disease severity.

SARS-CoV-2 lineage-specific disease symptoms and disease severity in São Caetano do Sul city, Brazil

Carlos A. Prete Jr.*[1,2], **Flavia C. S. Sales*[3]**, Leandro Abade[4], Lewis F. Buss[2,3], Darlan S. Candido[2,3], Ingra M. Claro[2,3], Filipe R. R. Moreira[5], Erika R. Manuli[3], Ligia Capuani[6], Camila A. S. Maia[3], Beatriz A. Oliveira[3], Thais Coletti[3], Heuder G. O. Paiao[3], Silvia F. Costa[3], Maria C. M. Correa[3], Fabio E. Leal[7,8], Kris V. Parag[2], Vítor H. Nascimento[1], Nuno R. Faria[2,3,4], Ester C. Sabino[3]

carlos.prete@usp.br **@carlos.prete1** **@CaddeProject**

1. Dept. of Electronic Systems Engineering, University of São Paulo, Brazil; **2.** MRC Centre for Global Infectious Disease Analysis and Department of Infectious Disease Epidemiology, Imperial College London, UK; **3.** Instituto de Medicina Tropicaland Dept. Moléstias Infecciosas e Parasitárias, Faculdade de Medicina, Universidade de São Paulo, Brazil; **4.** Dept. of Zoology, University of Oxford, United Kingdom; **5.** Dept. de Genética, Instituto de Biologia, Universidade Federal do Rio de Janeiro; **6.** Dept. de Tecnologia da Informação, Modular Research System Ltda, Brazil; **7.** Faculdade de Medicina, Universidade de São Caetano do Sul, Brazil; **8.** Divisão de Pesquisa Clínica, Instituto Nacional de Câncer, Brazil;
* Authors contributed equally..

Introduction

- The São Caetano do Sul city, southeast Brazil, established a web-based platform to provide primary care to suspected COVID-19 patients, integrating clinical and demographic data with representative genome sequencing.
- This study describes lineage-specific spatiotemporal dynamics of infections, clinical symptoms, and disease severity during the first year of the epidemic in the municipality.

Methodology

- We selected and sequenced 879 PCR+ swab samples (8% of all reported cases) between April 2020 and April 2021. Samples were selected uniformly across neighbourhoods and epidemiological weeks, yielding a spatially and temporally representative set of sequences.
- We developed an algorithm to infer daily lineage-specific prevalence and symptoms based on a moving window with time-varying width, allowing inference of cumulative cases and symptom probability stratified by lineage using integrated data from the platform.

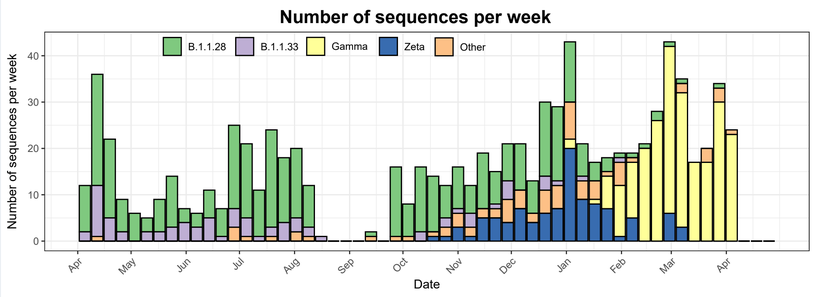

Fig 1. Number of sequenced PCR+ samples by epidemiological week identified as the main ancestral lineages (B.1.1.28 and B.1.1.33), VOI (Zeta) and VOC (Gamma). Samples not identified as one of these lineages were assigned into the group 'Other'.

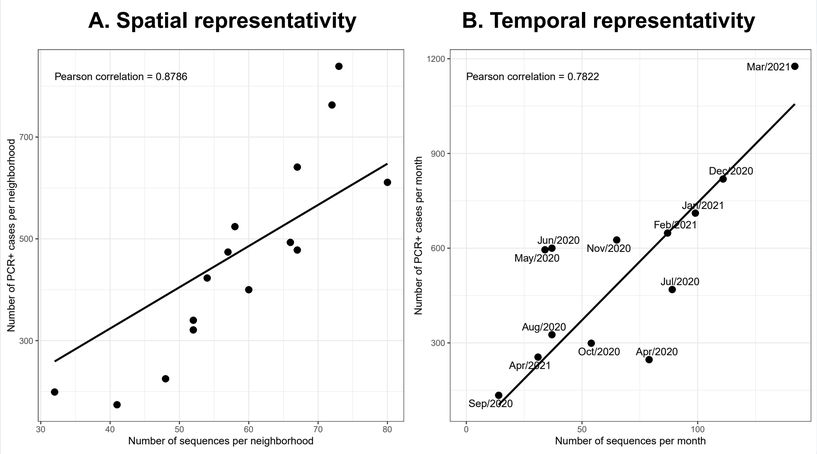

Fig 2. Correlation between number of PCR+ cases and number of sequences aggregated by neighbourhood (A) and month (B). Number of genomes strongly correlates with number of SARS-CoV-2 PCR+ confirmed cases both spatially and temporally.

Results

- Most infections were caused by B.1.1.28 (41.3%), followed by P.1/Gamma VOC (31.7%), P.2/Zeta VOI (9.6%) and B1.1.33 (9.0%).
- Gamma and Zeta were associated with larger prevalence of dyspnoea (respectively 81.3% and 78.5%) and persistent fever (84.7% and 61.1%) compared to B.1.1.28 and B.1.1.33.
- Ageusia, anosmia, and coryza were respectively 18.9%, 20.3% and 17.8% less commonly caused by Gamma, while altered mental status was 108.9% more common for Zeta.
- Effective reproduction numbers for B.1.1.28 and B.1.1.33 maintained an oscillating pattern around 1.0 due to persistent case importation from neighbouring cities.
- Case incidence was spatially heterogeneous and larger in poorer and younger districts.

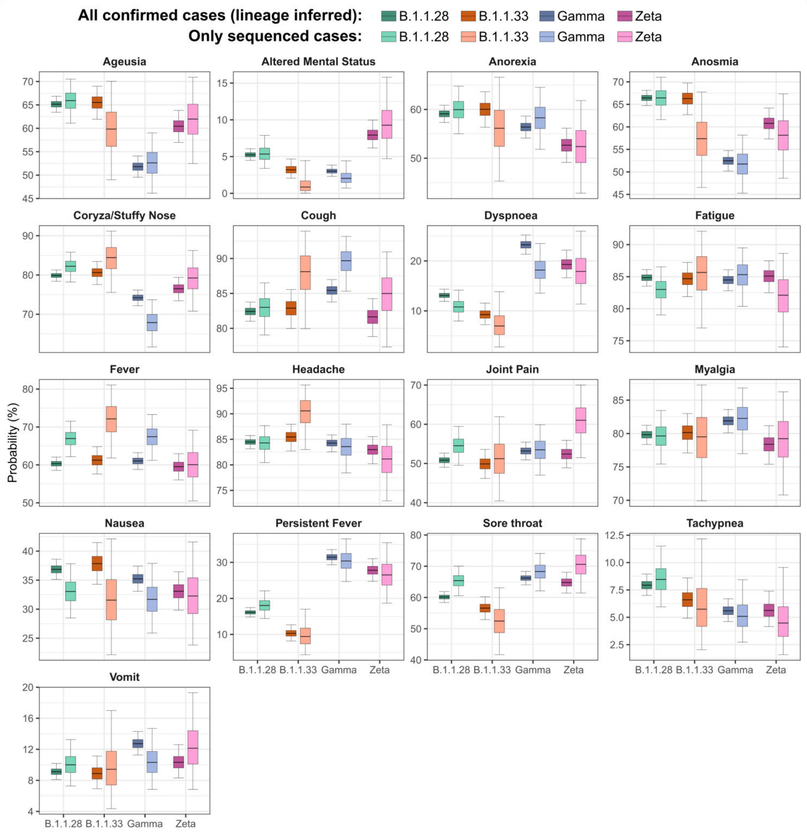

Fig 3. Probability of an individual infected with a given lineage reporting a specific symptom in any of the visits. Lighter boxes represent estimates obtained with only the 879 sequenced PCR+ cases, while darker boxes show probabilities inferred using all PCR+ cases. Lineage for non-sequenced PCR+ cases was imputed based on the prevalence of each lineage at the date of onset.

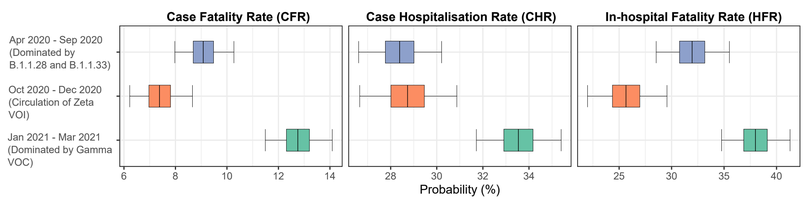

Fig 4. Inferred case fatality rate (CFR), hospitalisation rate (CHR) and in-hospital fatality rate (HFR) disaggregated by study period. All disease severity indicators were higher during the period when Gamma VOC was the most prevalent lineage.

Conclusion

- Our study demonstrates that Gamma VOC was associated with more severe disease, emphasising the role of its increased disease severity in the heightened mortality levels in Brazil.

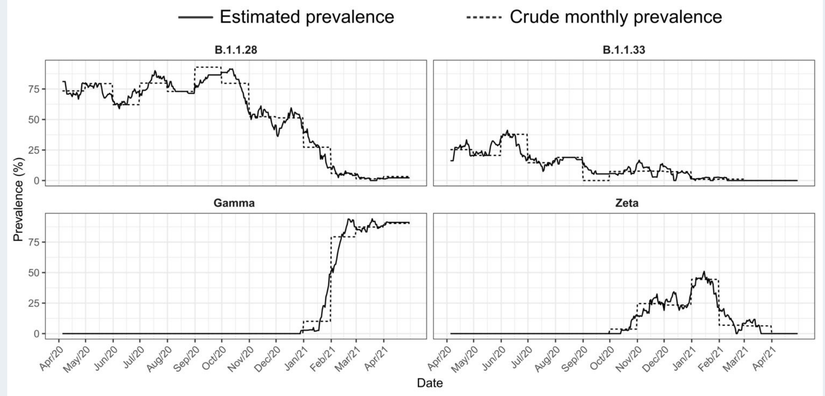

Fig 5. Estimated lineage prevalence compared with the crude monthly lineage prevalence, defined as the ratio between the number of samples of a given sequence in each month and the total number of samples in the same month.

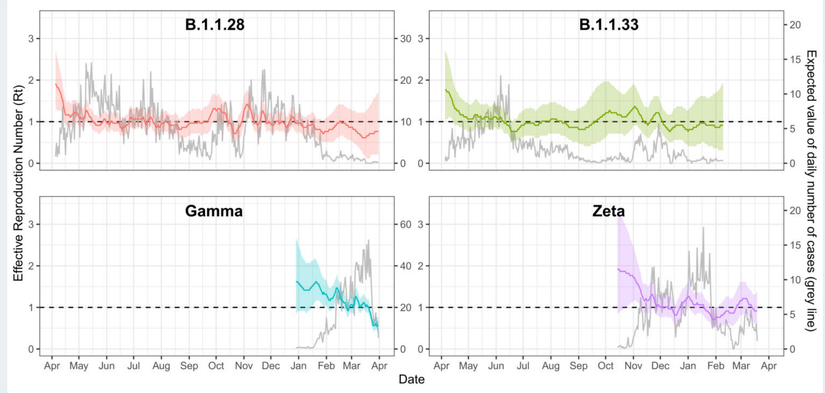

Fig 6. Lines show the median lineage-specific effective reproduction number (Rt) (y-axis on the left), and ribbons indicate 95% credible intervals. Grey curves represent the expected value of the estimated lineage-specific daily number of cases (y-axis on the right) by date of symptom onset.

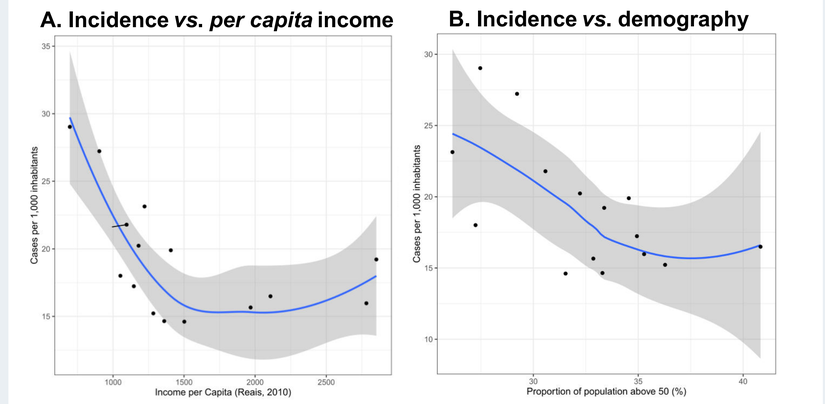

Fig 7. Total incidence in each neighbourhood compared to the income per capita (A) and proportion of population above 50 years old (B). The curve in blue and ribbons in grey respectively represent the mean and 95% confidence intervals of a Loess regression with span=1.0.

Acknowledgements

This work was supported by Bill & Melinda Gates Foundation (INV-034540 and INV-034652), MRC-FAPESP CADDE award (MR/S0195/1 and18/14389-0) (https://caddecentre.org) and Rede Corona-ômica BR MCTI/FINEP affiliated to RedeVírus/MCTI (FINEP 01.20.0029.000462/20, CNPq 404096/2020-4), FAPESP (2018/25468-9, 2022/15985-1, 2019/21858-0), Coordenação de Aperfeiçoamento de Pessoal de Nível Superior – Brasil (CAPES) – Finance Code 001, CNPq (308221/2022-2), Wellcome Trust and Royal Society Sir Henry Dale Fellowship 204311/Z/16/Z.

MRC Centre for Global Infectious Disease Analysis

Imperial College London

Poster

Preprint