20th Philippine Computing Science Congress (PCSC 2021),
Baguio City, Philippines

Computing Society of the Philippines

A Phylodynamic Analysis of the COVID-19 Spread in the National Capital Region of the Philippines

Imari C. Borda[1,2], Sheryl Grace Buenaventura[1,2], Zython Paul Lachica[1,2], Ivy Grace Panogalinog[2], Hanna Lyka Bontilao[1,2], May Anne Mata[1,2], Ricardo del Rosario[3]

[1] Center for Applied Modeling, Data Analytics, and Bioinformatics for Decision Support Systems, University of the Philippines Mindanao, Mintal, Tugbok District, Davao City, Philippines

[2] Department of Mathematics, Physics, and Computer Science, University of the Philippines Mindanao, Mintal, Tugbok District, Davao City, Philippines

[3] Broad Institute of MIT and Harvard, Cambridge, Massachusetts, United States of America

Abstract

To understand the disease dynamics in a particular location, incidence reports are used to estimate key epidemiological parameters such as transmission rates and reproductive numbers. However, incidence data often suffer from underreporting due to logistical concerns in disease surveillance, insufficient testing, etc. One way to address this concern is to use the phylodynamic approach – a method that estimates the epidemiological parameters and reconstructs the evolutionary history of an infectious disease directly from the genomic sequences of the vector.

Here, we apply a phylodynamic analysis method to study the COVID-19 dynamics in the National Capital Region (NCR) of the Philippines. Nineteen publicly available SARS-CoV-2 whole-genome sequences from NCR, sampled from 3 April 2020 to 18 July 2020, were used. We apply the Bayesian inference method to jointly estimate the epidemiological parameters under the birth-death Susceptible-Infected-Recovered (BDSIR) model and the evolutionary history from the viral sequence data. We estimate the joint posterior distribution using Monte Carlo Markov Chain (MCMC) algorithm, specifically, the Metropolis-Hastings (MH) sampler.

To compare the BDSIR model with standard epidemiological models, we created an ODE-based SIR model using reported incidence data and performed parameter estimation methods to determine epidemiological parameters. The data we used for the BDSIR model was different from the ODE-based model. The estimated transmission rate using viral genomic sequences was 2.67E-6, which is two orders of magnitude greater than the estimated transmission rate using the ODE-based SIR model (3.15E-8). Consequently, the estimated basic reproduction number of the BDSIR (3.12) is also higher than the SIR (1.22). Our results emphasize the need to cautiously use reported incidences and to consider the use of phylodynamic analyses as an additional tool from which to base decisions in implementing community-wide interventions in limiting the spread of infectious diseases.

Introduction

Coronavirus disease 2019 (COVID-19)

- An infectious disease caused by the Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2).
- The first case was reported from Wuhan, China on 31 December 2019. However, it is believed that the actual first case has occurred back on 8 December 2019.
- In Philippines, the first reported case (Patient 1) was confirmed on 30 January 2020.
- The second reported case (Patient 2) was confirmed on 2 February 2020. Patient 2 was a companion of Patient 1 during her travel from Wuhan to Manila.

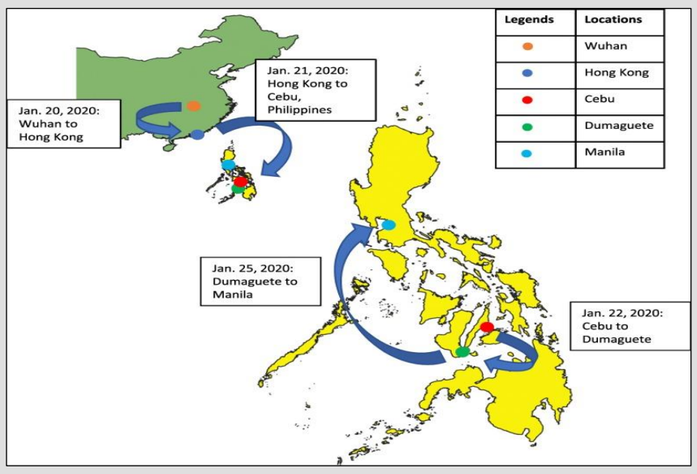


Fig 1. Travel history of Philippine's Patient 1 and Patient 2. They traveled from Wuhan on 20 January 2020 and arrived in Cebu on 22 January 2020 via Hong Kong. On 25 January 2020, they arrived in Manila from Cebu.

- Since the first case, the virus continued to spread in the Philippines (Fig 2). The first local transmission was confirmed on 7 March 2020.

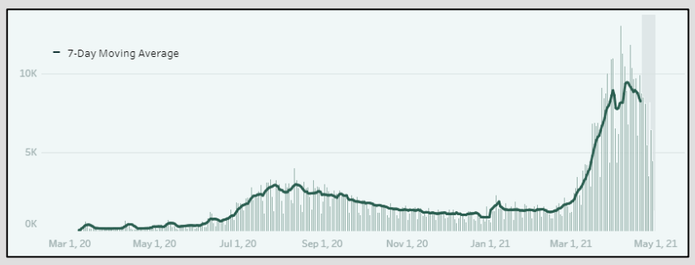


Fig 2. Philippine daily incidence of COVID-19 lifted from the Department of Health from 1 March 2020 to 1 May 2021.

- Common approaches in characterizing viral spreads rely on the reported incidences. However, incidence data often suffer from underreporting.

Main Objective

Main Objective: To understand the COVID-19 dynamics in the NCR Philippines by estimating the key epidemiological parameters directly from the genomic data of SARS-CoV-2 using phylodynamics.

Methodology

Data Gathering

- 19 SARS-CoV-2 whole-genome sequences from NCR sampled from 03 April 2020 to 18 July 2020.

National Center for Biotechnology Information (NCBI) **19** SARS-CoV-2 whole genome sequences from NCR

Phylodynamic Model: Birth-Death Susceptible-Infected-Recovered (BDSIR)

- First coined by Grenfell in 2004.
- Allows estimation of epidemiological parameters using only the genomic data of the vector.

Phylodynamic = Genomic + Epidemiological Model

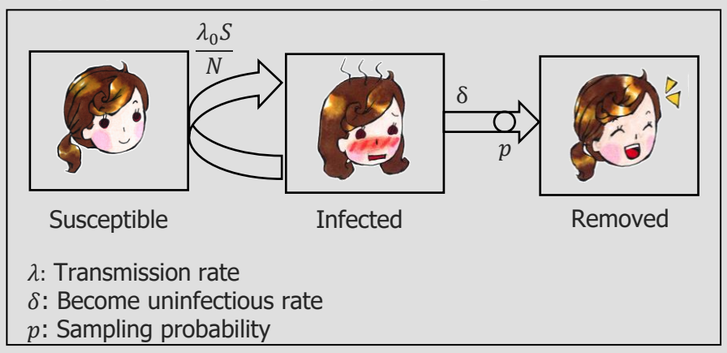


Fig 3. BDSIR model for COVID-19 transmission.

Bayesian Phylodynamic Inference

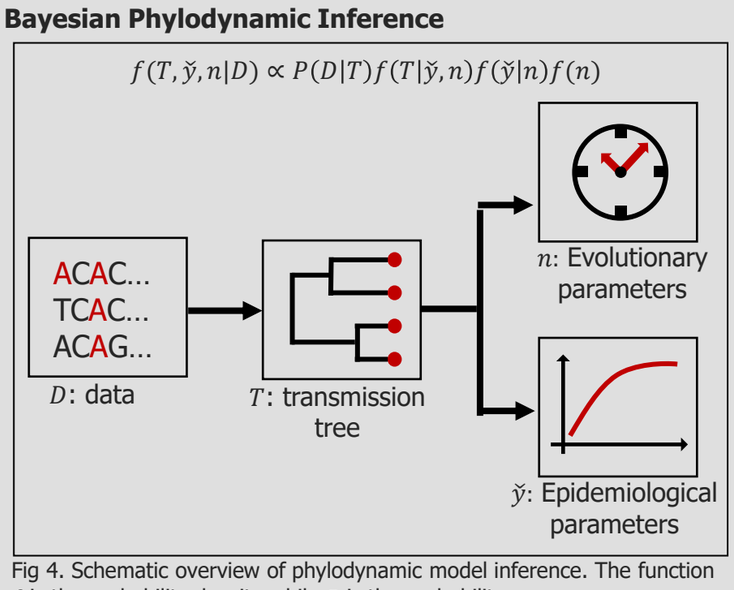
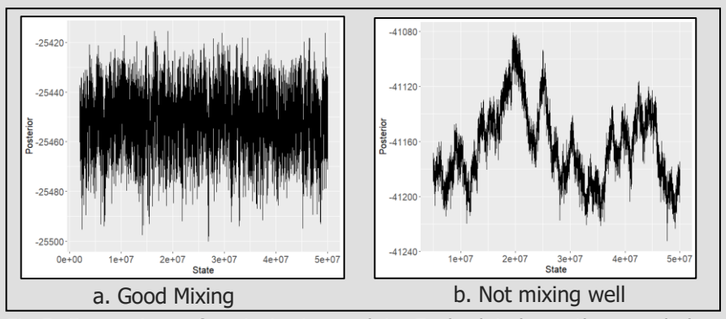


Fig 4. Schematic overview of phylodynamic model inference. The function f is the probability density while P is the probability mass.

Markov-Chain Monte Carlo (MCMC)

- MCMC is a stochastic algorithm that estimates the posterior distribution by performing a random walk.
- The Metropolis-Hastings algorithm:
  - Step 1: Get a starting parameter value, x_t.
  - Step 2: Propose a new parameter x' using the proposal density distribution.
  - Step 3: Calculate the acceptance ratio, α.
  - Step 4: Accept or reject.
    - If $\alpha \geq 1$, accept the new parameter.
    - Else, the new parameter is accepted with probability α.
  - Step 5: Repeat steps 2-5 a large number of times.
- Convergence of MCMC is assessed based on the ESS (Effective Sample Size) value and the trace plot of the posterior. An ESS value of at least 200 and a trace plot with no significant long range fluctuation suggests good mixing. Good mixing indicates that MCMC provides a good estimate of the posterior distribution.



a. Good Mixing b. Not mixing well

Fig 5. Comparison of posterior trace plots. Each plot shows the sampled values of posterior over time. A trace plot with good mixing shows no significant fluctuations while a trace plot that is not well-mixed has evident fluctuations.

Main Results and Conclusion

Assessing the Convergence

- The ESS of the posterior is 312 (≥ 200).
- The posterior trace shows no significant long range fluctuation (Fig. 6).
- Hence, we can be confident that the MCMC provides good estimate of the posterior.

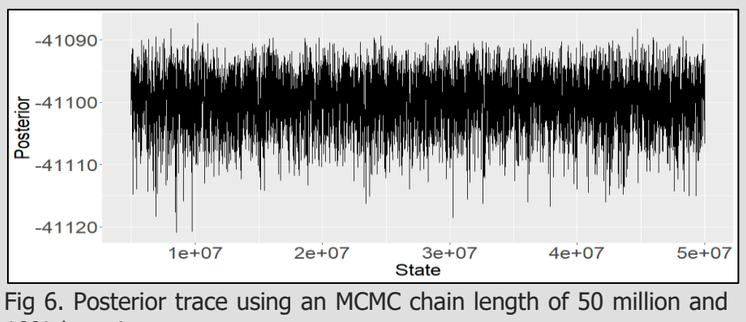


Fig 6. Posterior trace using an MCMC chain length of 50 million and 10% burn-in.

Parameter Estimates from BDSIR model and comparison with SIR model

- Using the 19 SARS-Cov-2 viral sequences from NCR, the BDSIR method estimated that the start of the pandemic was around 17 January 2020.
- The reproductive number estimate from BDSIR (3.12) was higher than the reported value (maximum value reported was 2.5) and was also higher than the estimate from the ODE-based SIR model using incidence data (1.22)
- The estimated transmission rate from BDSIR (2.67E-6) was also higher than the SIR estimate (3.15E-8).
- The incidence estimates from BDSIR were also higher than the SIR estimates.

Parameter	Median	95% HPD
Become Uninfectious Rate	7.12	[1.45, 30.46]
Birth (Transmission) Rate	2.67E-6	[1.25E-6, 5.61E-6]
Reproductive Number	3.12	[1.15, 7.56]

Table 1. Estimates of key epidemiological parameters from 19 NCR samples from 17 January 2020 to 18 July 2020, using the BDSIR model. HPD: Highest Posterior Density.

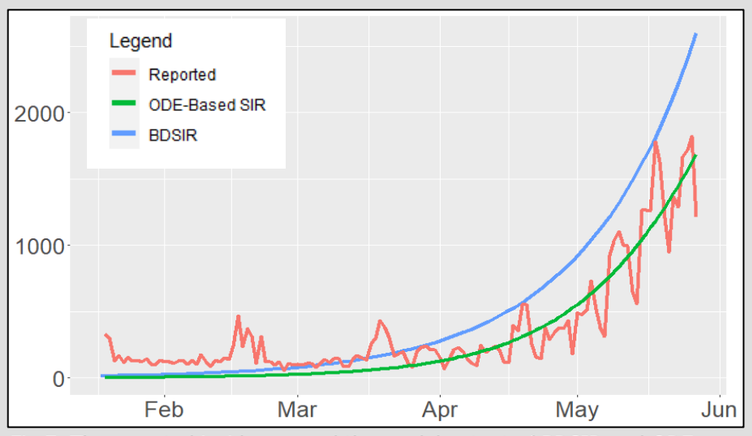


Fig 7. The reported incidence, and the model output of BDSIR and ODE-based SIR in NCR from 17 January 2020 to 18 July 2020.

Policy Implications

- Underreporting of epidemiological estimates is most likely to happen when underreported incidence data is used.
- Phylodynamic analysis should also be considered as an additional tool in characterizing COVID-19 spread and as one of the bases in crafting community-wide interventions.

Acknowledgement

Department of Science and Technology

Philippine Council for Health Research and Development

References

Kühnert, D., Wu, C. H., & Drummond, A. J. (2011). Phylogenetic and epidemic modeling of rapidly evolving infectious diseases. Infection, genetics and evolution, 11(8), 1825-1841.

Kühnert, D., Stadler, T., Vaughan, T. G., & Drummond, A. J. (2014). Simultaneous reconstruction of evolutionary history and epidemiological dynamics from viral sequences with the birth–death SIR model. Journal of the Royal Society Interface, 11(94), 20131106.

Contact Information

Center for Applied Modeling, Data Analytics, and Bioinformatics for Decision Support Systems
https://sites.google.com/up.edu.ph/nicer-upmin/home

University of the Philippines Mindanao, Mintal, Tugbok District, Davao City, Philippines

Phone: 0917 129 3986
Email: nicer.upmin@up.edu.ph