ARTICLE

Phylodynamic analysis of the spread of COVID-19 in the National Capital Region, Philippines during the early stages of the pandemic

Ivy Grace M. Panogalinog*[1,2,3,4,ϒ], Imari Joy C. Borda[1,ϒ], Zython Paul T. Lachica[1,2,3,4,§], Sheryl Grace C. Buenaventura[1], Ian Lorenzo Quibod[5,ʃ], Alexis Erich S. Almocera[1,3], Lyre Anni E. Murao[1,6], May Anne E. Mata[1,2,3,4], and Ricardo C.H. del Rosario[7]

[1]Center for Applied Modeling, Data Analytics, and Bioinformatics for Decision Support Systems in Health, University of the Philippines Mindanao, Mintal, Tugbok District, Davao City, 8022 Philippines
[2]Mindanao Center for Disease Watch and Analytics, University of the Philippines Mindanao, Mintal, Tugbok District, Davao City, 8022 Philippines
[3]Department of Mathematics, Physics, and Computer Science, University of the Philippines Mindanao, Mintal, Tugbok District, Davao City, 8022 Philippines
[4]University of the Philippines Resilience Institute, University of the Philippines Diliman, Quezon City, 1101 Philippines
[5]Rice Breeding Innovations Platform, International Rice Research Institute, DAPO Box 7777, Metro Manila, 4031 Philippines
[6]Department of Biological Sciences and Environmental Studies, University of the Philippines Mindanao, Mintal, Tugbok District, Davao City, 8022 Philippines
[7]Stanley Center for Psychiatric Research, Broad Institute of MIT and Harvard, Cambridge, Massachusetts 02142, United States of America

[§]Present Address: Nuffield Department of Primary Care Health Sciences, Radcliffe Primary Care Building, Radcliffe Observatory Quarter, Woodstock Rd, Oxford OX2 6GG, United Kingdom
[ʃ]Present Address: Plant Health Institute of Montpellier (PHIM), University of Montpellier, CIRAD, INRAE, Institut Agro, IRD, Montpellier, 34090 France

[ϒ]Equal contribution

ABSTRACT

In epidemiological studies, key parameters could be incorrectly estimated when case data are underreported, as demonstrated during the early spread of the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). Phylodynamic models, which integrate viral genomic data, offer an alternative approach to understand SARS-CoV-2 transmission dynamics. Here we aimed to understand the early spread of SARS-CoV-2 in the National Capital Region of the Philippines (NCR) through phylogenetic and phylodynamic analysis of viral sequences. We investigated how government-imposed community quarantines influenced viral spread and compared these estimates with those from conventional

*Corresponding authors
Email Address: **memata@up.edu.ph; rcdelros@broadinstitute.org**
Date received: 18 March 2026
Dates revised: 16 July 2026
Date accepted: 13 August 2026
DOI: https://doi.org/10.54645/2026192YRQ-29

KEYWORDS

Phylodynamic modeling; COVID-19; National Capital Region (NCR) Philippines; SARS-CoV-2 viral sequence analysis; epidemiological parameters

compartmental models. We analyzed 75 SARS-CoV-2 genomes sampled from NCR from 2020/03/13 to 2020/07/27. We used the Birth-Death Skyline (BDSKY) model to estimate the effective reproduction number (R_e), the Coalescent Skyline (COALSKY) model for the effective population size (N_e), and the Birth-Death Susceptible-Infected-Removed (BDSIR) model to estimate the basic reproduction number (R_0), transmission rate, and the number of susceptible, infected, and removed individuals. The estimated R_0 was 1.45 (95% Highest Posterior Density (HPD): 1.33–1.62). A major increase in R_e was inferred around 2020/05/15 (95% HPD: 04/26 to 05/30), coinciding with the easing of quarantine measures. Comparison of phylodynamic estimates and reported cases revealed that only 3% of cases were detected daily. Our findings, derived from a relatively small number of viral genomes, provide insights into the early spread of COVID-19 in NCR and demonstrate the utility of phylodynamic models for informing public health responses in the Philippines.

INTRODUCTION

The Coronavirus Disease 2019 (COVID-19) is an infectious disease caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). Described as a pneumonia-like disease, it was first reported in Wuhan, China, in late December 2019 (Zhu et al., 2020). The virus quickly spread to other countries and territories, leading the World Health Organization to declare it as a pandemic on 11 March 2020 (World Health Organization, 2020a). In July 2020, the pandemic had more than 24 million cases and over 800,000 deaths worldwide, disrupting the essential health services of 90% of countries, with low- and middle-income countries, including the Philippines, facing more significant challenges (Bong et al., 2020; World Health Organization, 2020b; Logrosa et al., 2022).

The Philippines reported its first confirmed COVID-19 case on 30 January 2020 from a 39-year-old female tourist from China who was admitted to the San Lazaro Hospital in Manila. Six months later, the COVID-19 Philippine statistics had reached more than 80,000 cumulative cases and almost 2,000 deaths (World Health Organization, 2020b; World Health Organization, 2020c). As the country's center of trade and commerce, Metro Manila, formally known as the National Capital Region (NCR; the largest urbanized area in the Philippines consisting of 16 cities), was considered to be the COVID-19 epicenter, which in 2020 accounted for 55% of cases and 52% of deaths in the country (World Health Organization, 2020c), and it accounted for 12.4% of the Philippine population as of 2020 (Philippine Statistics Authority, 2021).

Traditional epidemiological surveillance methods rely on reported data (i.e., line list of cases, deaths, and/or recovery) to provide estimates for epidemiological measurements and to model the disease dynamics. However, these methods are subject to inherent limitations arising from multiple factors, mostly related to a health system's ability to capture, trace, and properly diagnose symptomatic and asymptomatic cases (Arcede et al., 2022). Underestimation of the actual number of cases could result in an erroneous perception of the real extent of the virus spread. One way to address these limitations is through phylodynamic analysis, a method that utilizes molecular data from viral genomic sequences and integrates them with disease surveillance (Grenfell et al., 2004; Volz et al., 2013).

Viral genome sequences are an important secondary source of information on outbreak dynamics. The measurable evolutionary changes in viral genomes between sampling times can provide information about past transmission dynamics (Rambaut et al., 2016). Phylodynamic models based on Bayesian inference jointly estimate the epidemiological parameters and phylogenetic history from genomic data (Volz et al., 2013; Kühnert et al., 2014) and have been widely used to study the spread of COVID-19 (Rojas-Gallardo et al., 2020; Seemann et al., 2020; Danesh et al., 2021; Nadeau et al., 2021; Zhu et al., 2021). However, many of these studies have focused on countries in North America, Western Europe, and East Asia (Lu et al., 2020; Worobey et al., 2020; Du Plessis et al., 2021). Although a large-scale phylodynamic and phylogeographic analysis reconstructed the spread of SARS-CoV-2 in Southeast Asia, including the Philippines, it did not provide details of patterns of epidemics at the country or regional level (Zhu et al., 2021). There is a need for a study on the early spread of COVID-19 in the Philippines where the limited testing capacity resulted in a significant underreporting of cases (Esguerra, 2020).

In this paper, we focused on the early COVID-19 spread in the National Capital Region, Philippines from January to July 2020 because it had the highest reported cases in the country during the early phase of the epidemic. Additionally, the early phase had the highest number of high-quality SARS-CoV-2 sequences made publicly available as of the date the data were gathered for this study (17 July 2022). The early phase of the pandemic is a critical stage in providing insights into the future dynamics of COVID-19 spread as well as other possible influenza-like diseases (Arcede et al., 2022; Yap et al., 2023). Our analyses focused on the estimation of epidemiological parameters including the date of origin of the epidemic, the basic reproduction number (R_0), which represents the average number of secondary infections generated by a single infected individual in a susceptible population, the effective reproduction number (R_e), which represents the average number of secondary infections generated by an infected individual under current population conditions (Lim et al., 2020), and the effective population size (N_e), a measure of the number of infections contributing to the genetic diversity of the virus over time (Kanaka et al., 2023). These epidemiological parameters provided valuable insights into when the virus was first introduced, how fast it was spreading, and how the community quarantines affected its spread. Finally, we compared the epidemiological estimates obtained from the genomic sequences (using phylodynamic modeling) to the estimates obtained from the reported cases using conventional compartmental models and differential equation-based models.

METHODS

All computer scripts used to analyze the data are available in GitHub (Borda et al., 2024).

Alignment of viral sequences and creation of phylogenetic tree

We downloaded 122 publicly available SARS-CoV-2 whole-genome sequences sampled from patients who tested positive in NCR during the first seven months of 2020 and whose metadata includes year, month, and day of the sampling. These sequences were made available as of 17 July 2022 by the Global Initiative on Sharing All Influenza Data (GISAID; Shu and McCauley, 2017) and the National Center for Biotechnology Information (NCBI; Sayers et al., 2021), and were assembled and deposited by three groups: the Philippine Genome Center (PGC) (Tablizo et al., 2020), by researchers in Thailand (Velasco et al., 2020), and by the Research Institute for Tropical Medicine (RITM, 2021). The genome sequences were generated from nasopharyngeal and/or oropharyngeal swab specimens, and those generated by the PGC were sequenced using the Illumina NextSeq platform, whereas those generated by the RITM and reported by Velasco et al. (2020) were sequenced using the Illumina MiSeq platform. Consensus genome sequences were assembled using the ncov2019-artic-nf pipeline and associated tools, including BWA, SAMtools, and iVar (Velasco et al., 2020; Tablizo et al., 2020; RITM, 2021). Following the quality-control recommendations implemented in the Nextstrain quality control tool and SARS-CoV-2 Workflow (Aksamentov et al., 2021), we discarded low-quality genome sequences defined as those with more than 3,000 missing or ambiguous sites (approximately 10% of the SARS-CoV-2 genome). 75 sequences passed quality filtering and were

subsequently used for phylogenetic and phylodynamic analyses (see Table S1 for the list of 75 NCR sequences).

The 75 sequences were aligned and trimmed using the Nextclade command line interface, resulting in a final alignment of 30000 nucleotide sites. A time-scaled maximum-likelihood phylogenetic tree was created from this alignment using IQ-Tree v.2.2.0.2 and TreeTime. Ancestral state reconstruction was performed using the Augur software package incorporated in the Nextstrain platform (Hadfield et al., 2018). The Auspice tool was used to visualize and analyze the phylogenetic tree and other data in real-time (Hadfield et al., 2018).

A separate phylogenetic analysis using IQTree v.2.2.0.2 was performed to extract ultrafast bootstrap values, using 1,000 bootstrap replicates and using the General Time Reversible substitution model with empirical base frequencies (GTR+F+R2). The GTR+F+R2 was identified by the ModelFinder Plus tool as the best-fitting model according to the Bayesian information criterion (BIC) with 130,190.5. All other models had slightly higher BIC (range: 130,364.1 - 134,245.4 from 26 models tested in total (Borda et al., 2024)). The resulting maximum likelihood tree was visualized using FigTree V1.4.4 (Rambaut, 2018).

Bayesian Time-Scaled Phylodynamic Analysis using Bayesian Evolutionary Analysis Sampling Trees (BEAST)

The time-varying effective reproduction number (R_e) and the effective population size (N_e) were estimated using the Birth-Death Skyline (BDSKY) model (Stadler et al., 2013) and using the Coalescent Skyline (COALSKY) model (Drummond et al., 2005), respectively. Since different community quarantine measures were implemented in NCR during the collection period, in our estimation of R_e, we included a parameterization that reflected these community measures.

Both the COALSKY and BDSKY models were non-parametric models and did not assume any underlying host population structure. In the implementation of the BDSKY model, we used the tool TreeSlicer v.0.2.0 (du Plessis, 2018) to set the sampling proportion to zero before the first sampling date (13 March 2020) and to estimate the sampling proportions after the first sampling date. TreeSlicer was also used to simultaneously infer the date of the major change in R_e and to estimate R_e in conjunction with the implemented community quarantines. During the sampling period, the government implemented three quarantine measures in NCR: the Enhanced Community Quarantine (ECQ), the Modified Enhanced Community Quarantine (MECQ), and the General Community Quarantine (GCQ). The first measure, ECQ, was the most restrictive and was implemented on 17 March 2020 (see Table S2). In line with these measures, the R_e in the BDSKY model was estimated in a piecewise manner over four different time intervals, with the intervals bounded by dates that coincided with the implementation of the quarantine measures: 17 March 2020, 16 May 2020, and 01 June 2020. This allowed us to estimate the potential of the disease to be epidemic prior to the interventions and to observe the potential effect of the quarantine measures.

To generate the Bayesian Skyline plot (BSP) of the effective population size (N_e), we applied the COALSKY method to the 75 NCR viral genome sequences. This method uses a coalescent framework to infer N_e even before the first sampling date and estimates its changes over time by reconstructing the relative genetic diversity of the viral population. This method divides the time from the present to the tree's root into intervals, estimating a distinct N_e for each segment. The segment endpoints align with coalescent events, and the segment size is determined by the number of events within it (Müller and du Plessis, 2024).

To estimate the date of origin of the epidemic (this date of origin precedes the spread in the Philippines) and the basic reproduction number (R_0) of SARS-CoV-2 in NCR, we fitted a Birth-Death Susceptible-Infected-Removed (BDSIR) model described in Kühnert et al. (2014) to the 75 NCR viral genomic sequences. In contrast to COALSKY and BDSKY which do not assume an underlying host population structure, BDSIR assumes a Susceptible-Infected-Removed (SIR) structure of host dynamics. Similar to standard SIR epidemiological models, the BDSIR model also assumes that the underlying population structure involves hosts that can be divided into three mutually exclusive compartments: susceptible (S), infected (I), and removed (R). In terms of reaction kinetics, the stochastic SIR model used for the BDSIR model was parameterized as follows:

$$I + S \xrightarrow{\beta} 2I \quad (1)$$

$$I \xrightarrow{(1-s)\gamma} R_h \quad (2)$$

$$I \xrightarrow{s\gamma} R_s \quad (3)$$

In Equations (1), (2), and (3), the removed compartment R was split between the observed recoveries R_S (i.e., sampled) and the unobserved recoveries R_h (i.e., not sampled or hidden). This separation was explicitly parameterized in terms of the parameter sampling proportion s. Moreover, β is the mass-action infection rate and γ is the removal rate.

All three phylodynamic models (BDSIR, BDSKY, COALSKY) were implemented in Bayesian Evolutionary Analysis Sampling Trees (BEAST) v2.6.6 (Bouckaert et al., 2014) using a random tree as the starting phylogenetic tree. We used the BDSIR and BDSKY implemented in the *phylodynamics* and *bdsky* packages, respectively, while the COALSKY model we used was available in the core of BEAST software.

Initial Parameter Estimates

BEAST requires a number of input parameters, which are provided either as constant values or as prior distributions. The parameters used in this study are provided in Table S3 along with a brief justification for each. The clock rate is one such parameter that can either be set to a previously chosen value or be estimated from data. Since the number of sequences for the period of interest was limited, we fixed the clock rate to 8×10^{-4} substitutions/site/year to reduce the number of parameters to be estimated, which should result in better convergence when estimating the other unknown parameters (Drummond and Bouckaert, 2015).

The next parameter is the initial number of susceptible individuals which must be provided with a prior. In 2020, the population in NCR was estimated to be around 14 million (Philippine Statistics Authority, 2021). By 27 July 2020, i.e., the collection date of the last analyzed sample, community restrictions had already been implemented to limit the movement of individuals aged under 21 years or above 60 years old (Inter-Agency Task Force, 2020a). It was also projected that by the end of 2020, approximately 50% of the population will be between 21 to 60 years old (Philippine Statistics Authority, n.d.). Thus, for the initial number of susceptible individuals, we used a median prior estimate of 7 million (i.e., 50% of 14 million). Although this represents a reasonable approximation of population at risk, the influence of this assumption on the BDSIR parameter estimates was not formally evaluated. Thus, the inferred epidemiological parameters should be interpreted in light of this modeling assumption.

Another parameter is the prior for sampling proportion. By 27 July 2020, the total cumulative reported cases in NCR were around 44,491 (Department of Health, 2021). Since 75 genomic samples were used for this analysis, the prior for sampling proportion was set to have an upper bound of 1.69×10^{-3} (i.e., 75/44,491). The median of the becoming uninfectious rate (this parameter will be denoted "*become uninfectious rate*") was 20 days based on the estimate of Haw et al. (2020), with a standard deviation of four to account for the uncertainty. Note that the duration of the illness is

the reciprocal of *become uninfectious rate.* Also based on the work of Haw et al. (2020), we used a median prior estimate for the origin to be January 2020. Lastly, for the reproduction numbers, the prior was specified such that its 95% credible interval encompassed the range of estimated time-varying reproduction numbers based on the COVID-19 dashboard of the University of the Philippines (UP COVID-19 Pandemic Response Team, 2020) from March 2020 to July 2020 in Metro Manila. More details are provided in Table S3.

For the Monte Carlo Markov Chain (MCMC) used by BEAST, we used 50 million MCMC steps and sampled every 5000th step to reduce autocorrelation between MCMC samples (Drummond and Bouckaert, 2015). The first 10% of samples from the chain were discarded as burn-in (i.e., the removed initial part of the run where the MCMC is still in a low probability region of the parameter space). Convergence was assessed using Tracer v1.7.1 (Rambaut et al., 2018) by observing the effective sample size (ESS) and observing the marginal plots of the parameters. All parameters with an ESS of at least 200 and a marginal plot with no notable long-range fluctuations were considered to have converged (Drummond and Bouckaert, 2015).

We also compared the prior (i.e., no data) and posterior (i.e., with data) distributions of the parameter estimates to investigate whether the data (i.e., genomic sequences) was informative. We extracted the prior distributions by executing another independent MCMC run but sampling only from the priors. A parameter with a posterior that is significantly different from its prior means that the data was informative in estimating it (Drummond and Bouckaert, 2015). We statistically tested the difference between the two distributions by performing the Kolmogorov-Smirnov test in *R* and used a significance level of 0.01. This non-parametric test evaluates whether two sets of samples are drawn from the same underlying probability distribution (Birnbaum and Tingey, 1951; Marsaglia et al., 2003). If the resulting *P*-value is less than the level of significance (0.01), then there is sufficient evidence that the prior and the posterior distributions are significantly different.

Comparison of estimates to traditional models

Two traditional methods were used to model the COVID-19 cases using reported data: the deterministic SIR model and the stochastic SIR model. The ordinary differential equations (ODE) for the deterministic SIR model are shown below.

$$ds/dt = -\beta SI/N \quad (4)$$
$$dI/dt = (\beta SI/N) - \gamma I \quad (5)$$
$$dR/dt = \gamma I \quad (6)$$

Similar to the BDSIR model, S denotes the susceptibles, I the infected, and R the removed. The β is the effective transmission rate while γ is the removal rate. We assumed a closed population where the total population was constant at N. The model was fitted to the reported data in NCR from 1 March 2020 to 27 July 2020 obtained from the Department of Health (DOH) COVID-19 tracker as of 03 March 2022 (Department of Health, 2021). We first performed parameter estimation by numerically solving the ODE using MATLAB's *ode15s* for a given set of values of β and γ. For each candidate parameter set, we evaluated the numerical solution at the time points when cases were reported and calculated the mean squared error between the numerical solution and the reported cases. We then solved the optimization problem using MATLAB's *fminsearch* (MathWorks Inc, 2010). However, the optimization algorithms failed to converge to a stable solution, likely because multiple combinations of β and γ produced similar objective function values, resulting in a relatively flat optimization landscape (we have plotted the objective function and found the surface to have a relatively large flat area; data not shown) and sensitivity to the initial parameter values. We therefore performed a simple grid search. We constructed a 1000 by 1000 grid over possible values of parameters β and γ (between 0 and 4 for both parameters; the limits were determined heuristically). For each combination of β and γ in the grid, we solved the ODE over the time interval using *ode15s*, evaluated the ODE solution at the time points where the cases were reported, and calculated the mean squared error between the ODE solution and reported cases. The combination of β and γ with the lowest error was taken as the optimal parameter set. In all ODE simulations, we used 7,000,000 as the initial susceptible population, the average of the first five reported infected cases for the initial value of the infected population, and 0 as the initial removed population.

The kinetic reactions used in the stochastic SIR model are given by

$$S + I \xrightarrow{\frac{\beta}{N}} 2I \quad (7)$$
$$I \xrightarrow{\gamma} R \quad (8)$$

These stochastic reactions are equivalent to the ODEs in (4) – (6). To obtain the stochastic SIR trajectories, we used the best-fitting β and γ from the deterministic SIR and generated the trajectories using the Gillespie algorithm in the *pomp* package in *R* (King et al., 2016; Gillespie, 1977). We generated 100 realizations and calculated the median trajectory to represent the central tendency of the stochastic simulations, as the median is less sensitive to skewness and extreme values that may arise from stochastic variation.

The trajectories from the deterministic SIR and the stochastic SIR using the reported data were compared to the trajectories of the BDSIR model which uses genomic sequences. We also compared the basic reproductive number (R_0) estimates. Aside from the best fitting R_0, we also took the nine R_0 from the deterministic SIR model with the least error and compared it to the 95% Highest Posterior Density (HPD) interval from the BDSIR model.

We also estimated the average percentage of the reported cases by taking the average of the daily percent of reported cases from 1 March 2020 to 27 July 2020. The daily percent of reported cases was calculated by dividing the reported cases by the estimated number of infected individuals from the BDSIR model using genomic sequences.

RESULTS

Phylogenetic analysis of the 75 NCR samples

We downloaded 75 high quality viral genome sequences from the GISAID (Shu and McCauley, 2017) and the NCBI (Sayers et al., 2021) sampled from NCR between 13 March to 27 July 2020 that were available as of 17 July 2022 (Fig S1; the earliest collected sample on 13 March 2020 was one and a half months after the first reported case on 30 January 2020), and performed phylogenetic analysis. The genetic relatedness of the variants in each clade was supported by high ultrafast bootstrap values at key nodes in Fig. 1, where ultrafast bootstrap values greater than 70% were considered good support for the branches of the maximum likelihood phylogenetic tree (Minh et al., 2013).

All 75 samples collected during the early COVID-19 spread in NCR belong to three clades that have been previously identified by the Nextstrain method: clade 19A which first emerged during mid-January 2020, clade 20A which first emerged in mid-February 2020, and clade 20B which first emerged in mid-March 2020 (Fig. 1). Of the 75 samples, 18 (24%) belong to clade 19A (orange branches in Fig. 1), 10 (13%) belong to clade 20A (blue branches in Fig. 1), while the majority (47; 63%) belong to clade 20B (green branches in Fig. 1). The finding that the earliest clade detected in NCR was clade 19A is not surprising, given that clade 19A first emerged in Wuhan, China and played a significant role in the early global spread of the virus. Furthermore, given the geographic proximity between China and the Philippines, the detection of clade 19A in the Philippines soon after the start of the pandemic is

not surprising (Hadfield et al., 2018; Amul et al., 2022). Clade 20A originated from 19A and subsequently spread globally (Hadfield et al., 2018), followed by the emergence of Clade 20B from 20A (Hadfield et al., 2018). The detection of these three clades in the early stages of the pandemic in the Philippines reflects the rapid spread of the virus across nearby regions. No Variants of Concern (VOCs) were detected in this study, as VOCs had not yet been identified at that time.

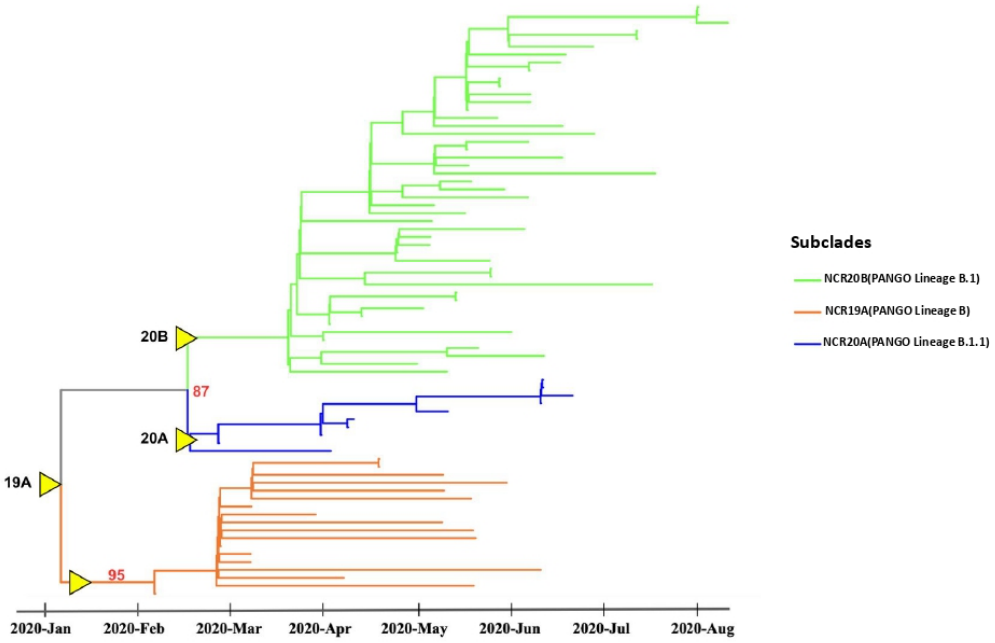

Figure 1: Time-scaled maximum likelihood phylogenetic tree. Tree reconstruction using the NextStrain algorithm from whole genome sequences of SARS-CoV-2 samples in the National Capital Region (NCR; n=75). The tree shows the hierarchical relationships among clades. Yellow triangles mark nodes indicating major Nextstrain clades and the colors of the branches indicate subclades: NCR/20B(PANGO Lineage B.1) (colored green), NCR/20A(PANGO Lineage B) (colored blue), and NCR/19A(PANGO Lineage B.1.1) (colored orange). Numbers in red indicate ultra-bootstrap values with more than 70% confidence.

To provide a global context for the phylogenetic relationships of the NCR samples, we chose 513 global sequences which were selected to have equal distribution across space (continents) and across time (in each month and year) and performed phylogenetic analysis on 588 samples (Fig. 2; see Table S4 for the list of sequences). The NCR sequences belonging to clade 19A also clustered together in the global phylogenetic tree, while a subset of clade 20B NCR sequences clustered together (Fig. 2, NCR/20B-1), indicating that these samples are from local transmission.

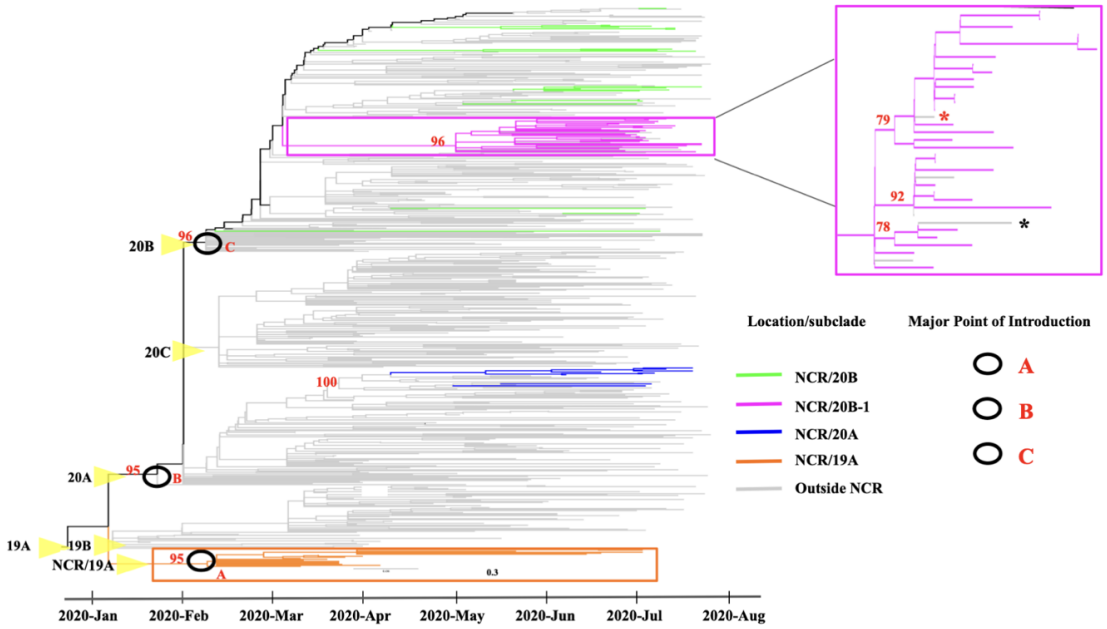

Figure 2: Phylogenetic analysis of NCR samples with global samples. Time-scaled maximum likelihood phylogenetic tree reconstructed using the NextStrain algorithm from whole genome SARS-CoV-2 samples in the National Capital Region (n=75) and from other regions inside and outside the Philippines (n=513). Yellow triangles: clades identified previously by Nextstrain. Gray branches: samples from other regions inside and outside the Philippines. Colored branches: NCR samples colored according to Nextstrain clades -- NCR/19A (orange; same samples as in Fig. 1), NCR/20A (blue; same samples as in Fig. 1), and the clade labeled NCR/20B in Fig. 1 is split into NCR20B (green) and NCR/20B-1 (magenta). Numbers in red: ultra-bootstrap values greater than 70% confidence (shown only for clades that contain NCR samples). Black circles labeled A, B and C: three major points of introduction of variants in NCR. Subclades NCR/19A and NCR/20B-1 (orange and magenta boxes, respectively) within clades 19A and 20B, respectively, trace geographic ancestry to NCR. Inset of NCR/20B-1: gray samples are from outside the Philippines; the two samples with asterisks are samples from Korea verified in GISAID to be imported cases from the Philippines (red asterisk) and with an unknown origin (black asterisk); gray samples without asterisks are samples from Hongkong.

Phylodynamic analysis reveals that effective reproduction number follows community quarantine measures

The effective reproduction number (R_e) is the expected number of secondary infections caused by an infected individual (Stadler et al., 2013). This quantity provides a measure of whether the epidemic is expected over time to die out ($R_e < 1$), grow ($R_e > 1$), or be steady ($R_e = 1$). The BDSKY model is a phylodynamic model used to infer the R_e of an epidemic from the viral genome data (Stadler et al., 2013). Here, we analyzed 75 NCR genomic sequences using three parameterizations of the BDSKY model to measure the R_e of COVID-19 in NCR: (1) parameterized for time-series estimation to measure the overall trend of R_e, (2) parameterized according to the community quarantines and to investigate how R_e behaves during quarantines and (3) parameterized for inferring the date of major change in R_e. Since BDSKY utilizes a coalescent framework, it permits inference of past transmission dynamics, allowing us to estimate R_e values before the earliest sampling date of the genomic sequences.

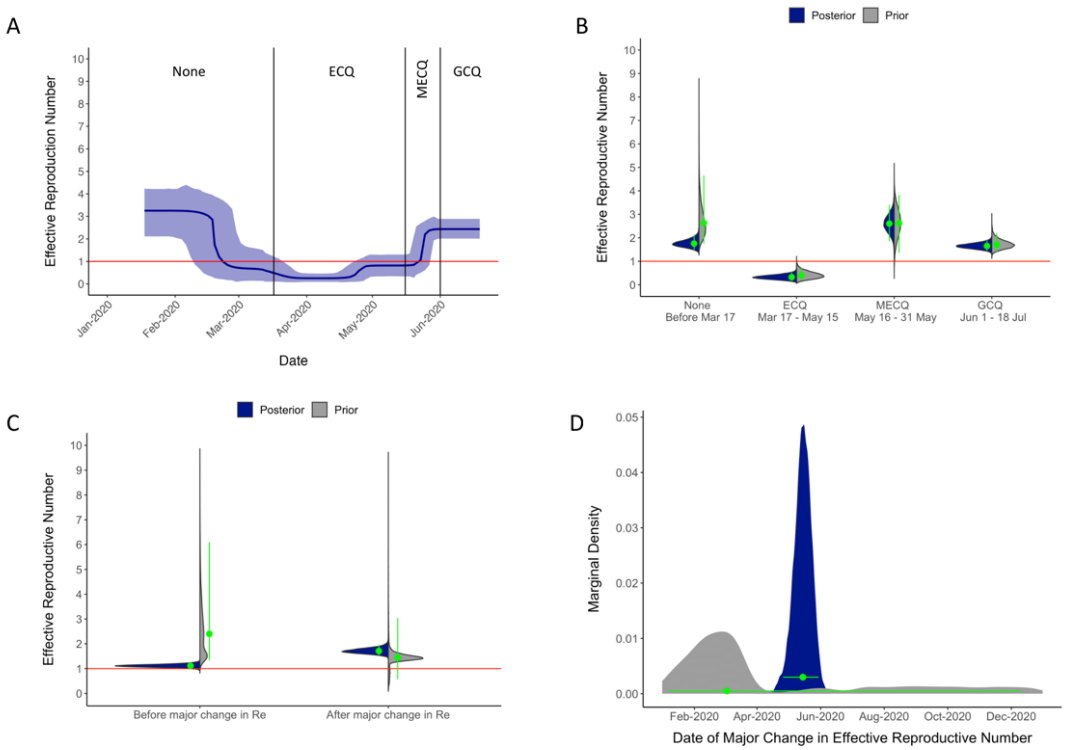

Figure 3: Effective reproduction number (R_e) estimates from the BDSKY model. (**A**) Smoothened time-series R_e estimates from BDSKY model over the analysis period (18 January 2020 to 27 July 2020; values in Table S5). The simulation start date of Jan 18 was from coalescent framework-based BDSKY estimates (191 days before the collection date of the first sample). Solid blue line: median value from 100 million MCMC chains; shaded blue area: 95% HPD interval. (**B**) Distribution of R_e estimates using a parameterization that reflects four community quarantine measures implemented by the government during the analysis period. Blue distribution: posterior estimates of R_e on each quarantine measure interval; Gray distribution: prior estimates. For each pair of prior and posterior distributions, the difference is significant (Kolmogorov-Smirnov test P-value<0.01). (**C**) Distribution of R_e estimates before and after the estimated date of major change in R_e, and (**D**) marginal density of the date of major change (blue: posterior distribution; gray: prior distribution). Horizontal red lines in (A), (B) and (C): epidemiological threshold ($R_e = 1$). Green dots in (B), (C) and (D): median value; green lines: 95% HPD interval. Quarantine measures: No quarantine (none), ECQ, MECQ, GCQ; see Table S2. The distributions illustrate a marginal probability density, i.e. the height (x-axis in (B) and (C); y-axis in (D)) of the distributions represents the proportion of the data located there, and the mode represents the consensus of the estimated values (y-axis in (B) and (C), x-axis in (D)). The estimates were from 18 January 2020 (estimated date of origin) to 27 July 2020 (collection date of last sample).

From the time-series parametrizations of the BDSKY, we were able to reconstruct the effective reproduction number of the epidemic in NCR with 95% credible interval from 18 January 2020 to 27 July 2020. Our results showed that R_e was decreasing from February to March 2020 (Fig. 3A). No quarantine measure was imposed during this period and the decrease in R_e could be due to the changes in the behavior of the population in response to the increasing number of cases. R_e remained low during the Enhanced Community Quarantine (ECQ) restrictions and although it started to increase in the middle of April, it remained below one during ECQ. (see Table S2 for the timeline of quarantine measures). The increasing R_e starting in mid-May coincided with less restrictions during the Modified Enhanced Community Quarantine (MECQ), with values reaching above 2.0 (Fig. 3A). However, this temporal association does not necessarily imply that the quarantine measures alone caused the reduction in transmission, as other factors such as changes in testing, population behavior, and mobility may also have contributed.

Fig. 3B shows the R_e estimates parameterized with respect to the implemented community quarantine measures in NCR (No interventions/None, ECQ, MECQ, GCQ). At each quarantine period, the posterior distribution of R_e (Fig. 3B blue plots) is distinct from the prior distribution (Fig. 3B gray plots; Kolmogorov-Smirnov test for difference in two distributions P-value<0.01). In agreement with the time-series parameterization in Fig. 3A, the R_e estimates were less than 1 during the ECQ and greater than 1 during MECQ and GCQ (Fig. 3B).

BDSKY parameterization to infer the date of major change indicated a major increase in R_e on 15 May 2020 (Fig. 3D; 95% HPD: 26 April 2020 – 30 May 2020). During this transition date, R_e increased from 1.11 (Fig. 3C; 95% HPD: 1.01-1.2) to 1.71 (Fig. 3C; 95% HPD: 1.53-1.96). To verify this, we plotted the posterior and prior distributions of R_e, before and after the major transition, and found that posterior distributions have a more focused distribution compared to the prior distributions (Fig. 3C). Notably, the date of the major increase in R_e corresponds to the time when the long period of ECQ (around 2 months) was about to transition to the more relaxed MECQ. While this temporal correspondence is consistent with changes in transmission following the easing of restrictions, it should not be interpreted as direct evidence of causality because multiple epidemiological and behavioral factors may have influenced transmission during this period.

Effective population size estimates using phylodynamic analysis

Next, we used the COALSKY method on the 75 NCR viral genome sequences to determine the Bayesian skyline plot of the effective population size (N_e) (Fig. 4). The Bayesian skyline plot infers relative genetic diversity as the product of the effective population size and generation time, and N_e represents the number of unique genomes capable of producing progeny virions and indirectly reflects the number of infected individuals in the context of the epidemic. The Bayesian skyline plot analysis has provided insights into the relative genetic diversity of viral populations, and could help in understanding the dynamics of transmission and evolution during a pandemic (Giovanetti et al., 2021; Frost and Volz, 2010). Similar to the BDSKY method, the COALSKY method also uses a coalescent framework that allows us to infer N_e before the date of first sampling.

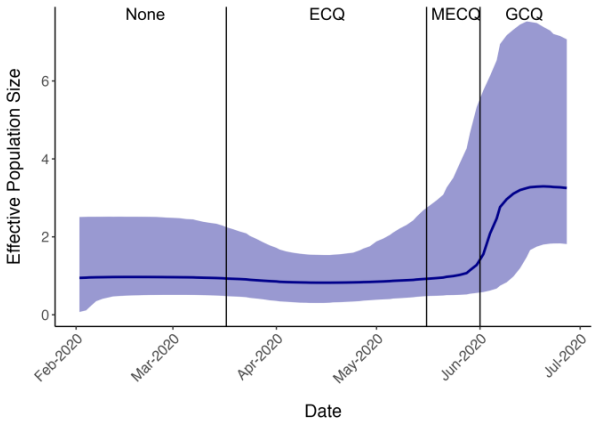

Figure 4: Effective population size (N_e) estimate from the COALSKY model. N_e estimates of the COALSKY model during the analysis period (31 January 2020 to 27 July 2020). Solid blue: median posterior estimate from 100 million MCMC chains; shaded blue area 95% HPD interval. Timeline of interventions: No quarantine, ECQ, MECQ, and GCQ.

The inferred N_e in Fig. 4 indicates that the population size of the virus was relatively constant during the first five months of 2020, suggesting a constant relative genetic diversity during this period. This stability may indicate a period of equilibrium in the population's evolutionary history, and also likely reflects a stable generation time given that quarantine measures have not significantly impacted the viral population dynamics. A population expansion took place from late May 2020 to early June 2020 (a population increase of two- to threefold). This is consistent with our BDSKY model which showed a major expansion of R_e around the same period. The increase in relative genetic diversity is likely due to the introduction of additional lineages and the accumulation of mutations within the viral population, which aligns with the relaxation of quarantine measures, specifically, GCQ during this time. Furthermore, we expect the generation time to have changed during this period due to the relaxation of quarantine measures, and this change may be reflected in the increase shown in the Bayesian skyline plot. This change was, however, not reflected in the reported cases in NCR (Fig. S2) (Department of Health, 2021).

Origin and transmission patterns in NCR

The BDSIR model is another phylodynamic model used to infer epidemiological parameters using viral genome sequence data (Kühnert et al., 2014). It is a parametric model and has an underlying compartmental model for the host population involving the susceptible, infected, and removed individuals. Here, we used the BDSIR model to estimate the four epidemiological parameters of COVID-19 based on 75 NCR genomic sequences: (1) the date of origin, (2) the basic reproduction number (R_0), which is the average number of secondary infections produced by one infected case in a population where everyone is susceptible, (3) the *become uninfectious rate* (rate at which infected individuals become un-infectious, recovered or dead), and (4) the transmission rate (Fig. 5). The model also allowed us to estimate the number of susceptible, infected, and removed individuals (Fig 6).

Our results indicate that the estimated onset of the global pandemic based on the NCR samples was around 23 December 2019 (Fig. 5A; 95% HPD: 16 November 2019 to 12 January 2020) with R_0 of 1.45 (Fig. 5B; 95% HPD: 1.33 to 1.62). Note that not all of the viral sequences were from local transmission (Fig. 1) hence the estimated onset is not the onset in NCR. The *become uninfectious rate* parameter was estimated at 38.7 individuals per year (Fig. 5C; 95% HPD: 32.4 to 45.2), suggesting that the infectious period of patients ranged from 8.08 (365/45.2) to 11.25 (365/32.4) days, with a median of 9.4 (365/38.66) days. The sampling proportion was 1.65×10^{-3} (Fig. 5D ; 95% HPD: 1.49×10^{-3} to 1.68×10^{-3}).

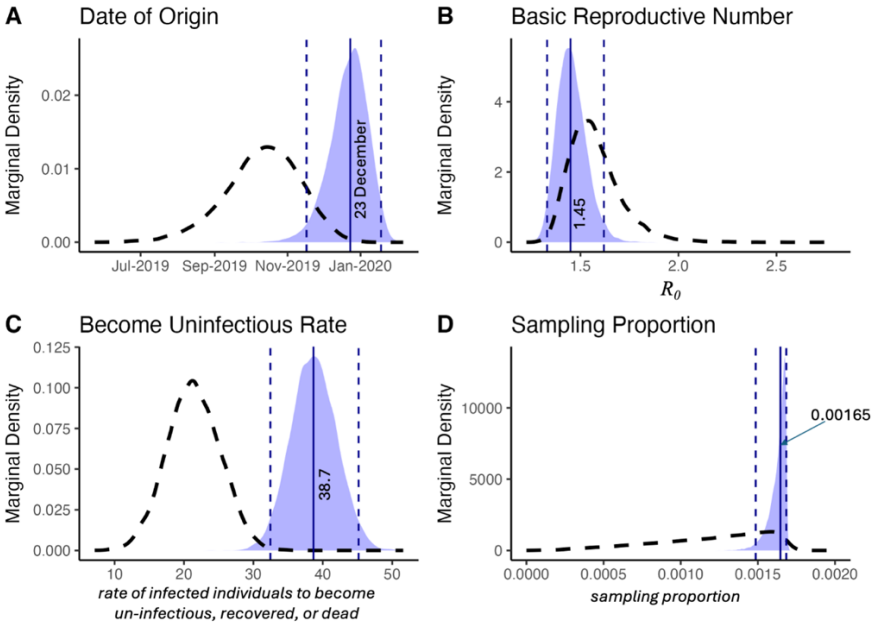

Figure 5: Comparison of posterior and prior distributions of key epidemiological parameters estimated from the BDSIR model. Blue shaded regions: posterior distributions, dotted black lines: prior distributions of estimates of (**A**) date of origin, (**B**) basic reproductive number, (**C**) *become uninfectious rate* (individuals/year), and (**D**) sampling proportion. Solid vertical blue lines: median of posterior distributions; dashed blue lines: lower and upper bounds of the 95% HPD interval of the posterior.

For both the date of origin (Fig. 5A) and the *become uninfectious rate* (Fig. 5C), the posterior distributions showed a shift to the right relative to their respective prior distributions. This implies that the duration of the COVID-19 illness was actually shorter and the date of origin may have been an earlier date than previously believed. The posterior distribution of R_0 showed a slight shift to the left relative to its prior distribution (Fig. 5B).

Among the four estimated parameters, R_0 showed the closest match between its posterior and prior distributions, which was likely due to the strong prior belief about its value. R_0 was the value frequently reported publicly which allowed us to provide a more informed prior distribution than the other parameters. On the other hand, a quasi-uninformative uniform prior distribution was used for the sampling proportion (Fig. 5D), but the resulting posterior was left-skewed due to sampling variability, which is a common issue in the analysis of rare events, specifically, the early spread of COVID-19 in NCR (Lam and Duchene, 2021; Frost et al., 2015). The estimates for the other three parameters, we believe, are less affected by the low number of sequences in different ways. The date of origin is influenced primarily by the phylogenetic structure of the data and the timing of the sampled sequences, which remains robust even with a small sample size. The basic reproduction number and *become uninfectious rate* are estimated through epidemiological modeling, which relies on data about the number of cases and transmission patterns and thus are relatively stable as they are informed by broader epidemiological trends.

Thus, although only 75 genomic sequences were used in our study, all four key epidemiological parameters (date of origin, R_0, *become uninfectious rate*, and sampling proportion) showed more precise posterior marginal distributions than the prior distributions (i.e., posterior distributions were taller and had a smaller variance in Fig. 5). This suggests that the data contributed substantially to estimating the parameters and that the confidence in the estimates increased when data were incorporated.

Figure 6 shows the estimated number of susceptible, infected, and removed individuals from 23 December 2019 (estimated median origin) to 27 July 2020. There was a steady increase in the number of infected individuals while the number of susceptible individuals showed an opposite trend. By the end of the sampling period, more than 60,000 individuals were estimated to be infected, and when we compared with reported data and numerical estimates using traditional SIR differential equation models, we found that only 3% of actual cases were reported (see next section).

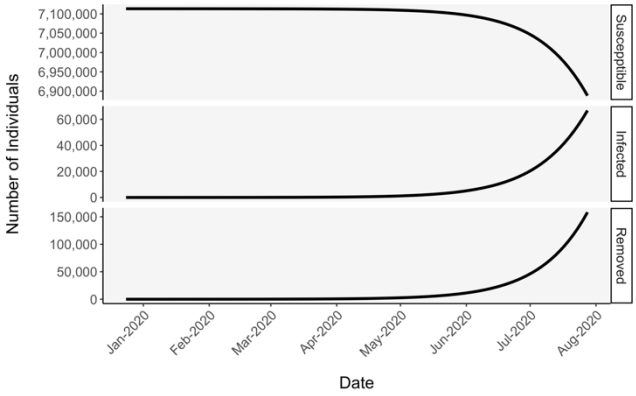

Figure 6: Estimates of the number of susceptible, infected, and removed using the BDSIR model. BDSIR results from 23 December 2019 (estimated date of origin) to 27 July 2020 (collection date of the last analyzed sample). Three compartments for the SIR model are shown: susceptible, infected, and removed.

Comparison of R_0 and trajectories between phylodynamic BDSIR model and differential equation-based SIR models

We next sought to compare the results of the BDSIR model to two differential equation-based compartmental models, the deterministic SIR and the stochastic SIR models. Similar to BDSIR, the SIR model has two parameters, the transmission rate β and the recovery rate γ (γ is also the reciprocal of the infectious period) (Methods; see equations (4)-(6) and (7)-(8), for the deterministic and stochastic models, respectively). For a given value of the parameters, we used numerical methods to solve the differential and stochastic differential equations and by naively fitting the models to the reported case data (Fig. S2), we calculated the best fitting values of β and γ.

The optimal values for β and γ from the deterministic SIR model were 0.09 and 0.06, respectively which is equivalent to an R_0 of 1.46 ($R_0 = \beta / \gamma$). The SIR-estimated R_0 agrees well with the estimate of the BDSIR model (R_0 = 1.45; 95% HPD from 1.33 to 1.62). However, it must be noted that the R_0 from the BDSIR model covered the period from 23 December 2019 (estimated median origin) to 27 July 2020 (collection date of the last analyzed sample), while the R_0 from the reported data covered a shorter time interval from 1 March 2020 to 27 July 2020, which was due to the availability of the reported data from the DOH of the Philippines. Thus, it is likely that the R_0 estimated from the SIR models would be lower if reported cases were available earlier, as in January 2020, since transmission was slow during that initial period.

A comparison of the number of infected individuals as estimated by the three models (BDSIR, deterministic SIR, and stochastic SIR) shows that, as expected, both the deterministic and the stochastic SIR trajectories closely followed each other, while the BDSIR model showed a higher number of infected individuals and a higher growth rate or slope (Fig. 7; BDSIR relative growth rate = 0.05, i.e., cases grow by about 5% of the current cases per day versus deterministic SIR relative growth rate = 0.03). The parameters estimated from deterministic SIR were used in the stochastic SIR simulations. By the end of the sampling period, the BDSIR estimated more than 60,000 infected cases while the actual reported data, and the estimates from traditional SIR model were only around 2,000 infected cases, which account for only 3% of the BDSIR estimates (2002/66781.5).

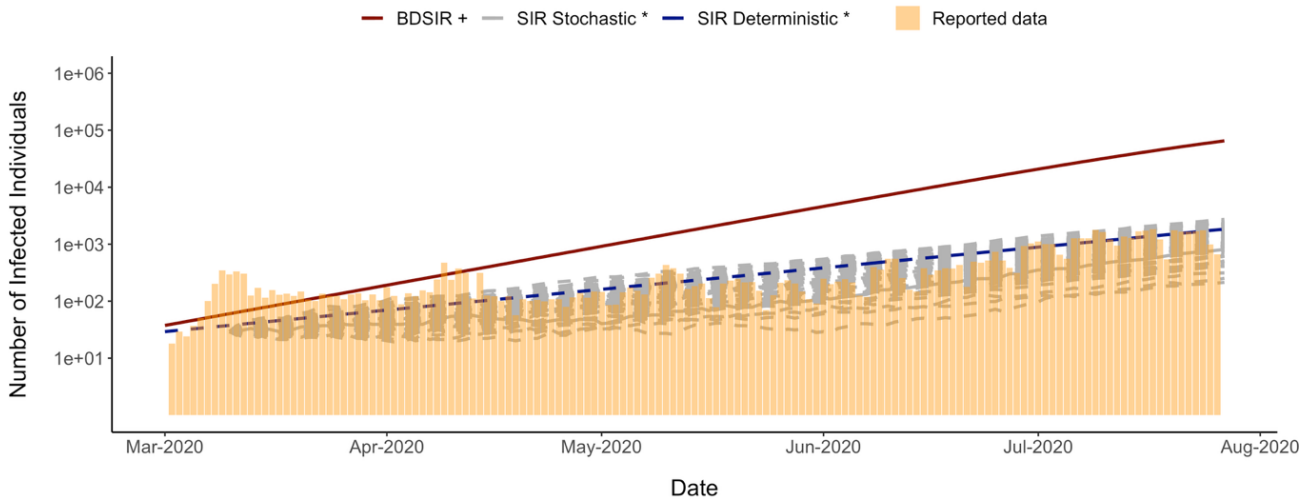

Figure 7: Comparison of the number of infected individuals between models using genomic data and reported case data. Red line: BDSIR model using viral genome data; blue dashed line: deterministic ordinary-differential equation-based model using reported case data; gray dashed lines: 100 realizations of the stochastic SIR model (using reported case data); orange vertical bars: reported cases from the DOH COVID-19 Tracker. *y*-axis: number of infected individuals in log scale. The symbol (+) in the legend denotes that the data used was genomic data while the symbol (*) denotes data used was reported case data.

In our study, we have demonstrated the disparity in the estimates of the infectious population between compartmental and phylodynamic models. Although fitting epidemiological compartmental models to reported data has been a common approach to estimating rates associated with disease dynamics, underreporting can significantly bias the estimates of disease prevalence and transmission rates leading to inaccurate predictions about the trajectory of an outbreak. Although some of these biases can be addressed by adding complexity to the compartment models, it is not always guaranteed to provide accurate estimates as it requires careful model calibration and precise parameterization. By contrast, phylodynamic models can incorporate genetic data from a subset of reported cases to infer the total number of infections and track the spread of the disease over time. Phylodynamic modeling has been shown to be particularly useful for diseases with low-case reporting rates (Volz et al., 2013), as in the case of the early pandemic in the Philippines, where testing laboratories were limited during that time (Amit et al., 2021).

CONCLUSION

Our phylogenetic analysis indicates that the onset of the global pandemic was estimated to be as early as in 2019 (i.e., estimated time to the most recent common ancestor (TMRCA)) (Hadfield et al., 2018)), which corresponds with the emergence globally of the ancestral clade 19A. However, the initial local transmission on 7 March 2020 might not have been fully captured by the phylogenetic analysis since all samples in the earliest clade NCR/19A came from the same geographic location, there was no apparent temporal clustering (Fig. 1). A possible explanation could be the low number of sequences (10 samples) at that time point (Department of Health, 2020).

Phylodynamic analysis estimated the date of origin of the global pandemic represented by the NCR samples to be around 23 December 2019, (HPD: 16 November 2019 to 12 January 2020), with the HPD interval not containing the reported first case (30 January 2020) (Edrada et al., 2020), suggesting a delay in the reporting of cases, likely due to surveillance complexities. The estimated median R_0 (1.45) inferred for the July 2020 period was higher than the DOH's estimate of 1.19 (World Health Organization, 2020d), potentially due to the under-reporting of cases in the DOH data. The infectious period was estimated from 8.08 to 11.25 days, which agreed with previous studies (10 days in mild-moderately ill patients and 15 days in severely-critically ill and immunocompromised patients (Rhee et al., 2021)). These results and those discussed below suggest that the phylodynamic framework can reasonably estimate the epidemic parameters in NCR.

The effective population size (N_e) estimates showed a slow increase from February to May 2020 (Fig. 4), which is slower compared to the initial growth observed from the Southeast Asian isolates, where N_e almost reached the value 10 by the end of May 2020 (Zhu et al., 2021). This could be due to inadequate sampling during the early pandemic stages. It should be noted that the oldest viral genome was obtained on 13 March 2020, 81 days after the estimated epidemic origin date of 23 December 2019, and only a small portion of the samples was collected before May 2020 – 12 out of 75, (16%). These limitations in early-period sampling likely resulted in an inadequate number of coalescent events to accurately estimate N_e during the early phase.

Both the BDSKY and COALSKY models showed an increase in positive cases in May (Fig. 3 for the former, Fig. 4 for the latter), which coincides with the period when the community quarantine changed from ECQ to MECQ (Inter-Agency Task Force, 2020b). During MECQ, restrictions were relaxed allowing for increased public gatherings, expanded mall and shopping center operations, and more open inbound travel for returning Overseas Filipino Workers, who constitute 5.67% of incoming international travelers (Inter-Agency Task Force, 2020a), which likely contributed to the increase in cases. Another possible explanation is a potential "adaptation period" following cross-species transmissions (Zhu et al., 2021), but this is unlikely considering that the inferred population growth already occurred more than three months after the first reported infection in NCR (Parrish et al., 2008). Pajaron et al. (2021) and Añonuevo et al. (2023) also observed a similar pattern in which COVID-19 cases spiked and transmission rates increased after the Philippines eased community quarantine restrictions.

The BDSIR model suggests that by 27 July 2020 (the last date of the sample), only 3% of estimated actual cases were included in the daily DOH reports (Department of Health, 2021). This underreporting is likely due to the low number of tests conducted in the Philippines (Hasell et al., 2020), with only an average of 13.2 tests per 1000 individuals and only 1.3% of the population being tested, as well as prevalent delays in the exact reporting of cases (World Health Organization, 2020e). Other studies have similarly

found low levels of COVID-19 case reporting, with estimates as low as 1-2% of the actual cases (Lau et al., 2021). This has also been reported in other countries; in Latin America, Rojas-Gallardo et al. (2020) found high numbers of unreported cases, ranging from 2 to 2200 times the actual reported numbers. This highlights the value of phylodynamic methodology in epidemic modeling, as it can potentially provide a more representative estimate of the number of cases (Frost and Volz, 2010). Additionally, this paper highlights the need to strengthen timely contact tracing efforts to monitor the spread of the virus.

Our study underscores the importance of genomic surveillance and timely contact tracing to better understand and control COVID-19 transmission. Moreover, a higher number of samples in future studies could better satisfy the model assumption of random sampling, since during the initial stages of the pandemic, sampling was based on who was identified as infected by contact tracers or health officers. As genomic surveillance in the Philippines continues to expand beyond the data available for this study, larger and more representative genome datasets will enable more comprehensive phylodynamic analyses and improve the precision of epidemiological parameter estimates. Although we used a fixed molecular clock with a clock rate derived from the Asian phylogenetic tree in Nextstrain (Hadfield et al., 2018), this assumption is expected to have minimal effect since the viruses in NCR belonged to the same clade as those from Asia. Our results provide an early snapshot of the transmission of the SARS-CoV-2 isolates circulating in NCR. We believe that the findings from this study will contribute to our understanding of COVID-19 epidemiological dynamics, and we expect that our results will be useful for parameterizing more specialized epidemiological models and informing future epidemic interventions at the regional or national level.

Our study is limited by the low number of high-quality viral genomes in NCR during the early stages of the pandemic. Although the sampled genome sequences spanned the study period, they likely do not fully represent all SARS-CoV-2 transmission occurring in NCR because sequencing was largely dependent on case detection and sample availability rather than systematic random sampling. This sampling bias may have influenced the inferred phylogenetic relationships and phylodynamic parameter estimates, particularly during the early months when only a limited number of genomes were available. Therefore, we were not able to perform an analysis from locally transmitted viral genomes. Nevertheless, we believe that our study underscores the importance of genomic surveillance in understanding the spread of the virus. Notably, our analysis aims to assess whether phylodynamics, despite relying on a small number of early samples, can provide better estimates of transmission during the early pandemic phase, prior to the emergence of distinct Variants of Concern.

ACKNOWLEDGMENT

This study was supported by the Department of Science and Technology (DOST) through the Predictive Modeling and Viral Phylodynamic Analysis on the Spatial and Temporal Patterns of Disease Outbreaks with considerations for Control and Logistics (PPASTOL) applied in Mindanao region research project under the Niche Center in the Regions for R&D (NICER) Program on Decision Support Systems in Health based in the University of the Philippines Mindanao and is monitored by Philippine Council for Health Research and Development (DOST-PCHRD). We thank the Department of Science and Technology – Computing and Archiving Research Environment Advanced Science and Technology Institute (DOST-COARE ASTI) for the HPC server used in computations.

CONFLICT OF INTEREST

The authors declare no competing interests.

DATA AVAILABILITY STATEMENT

All viral sequences analyzed in this study are available in the GISAID (Shu and McCauley, 2017) and NCBI repositories (Sayers et al., 2021). The scripts used to analyze the data are available at (Borda et al., 2024).

CONTRIBUTIONS OF INDIVIDUAL AUTHORS

I.G.M. Panogalinog: Conceptualization (supporting), Investigation (supporting), Methodology (supporting), Software (lead), Validation (supporting), Visualization (lead), Writing-review and editing (equal). **I.J.C. Borda:** Conceptualization (supporting), Data Curation (lead), Formal Analysis (lead), Investigation (lead), Methodology (lead), Software (lead), Validation (supporting), Visualization (lead), Writing-original draft (lead), Writing-review and editing (lead). **Z.P.T. Lachica:** Conceptualization (supporting), Formal Analysis (supporting), Funding Acquisition (supporting), Methodology (supporting), Project Administration (supporting), Validation (supporting), Writing-original draft (supporting), Writing-review and editing (supporting). **S.G.C. Buenaventura:** Conceptualization (supporting), Data Curation (supporting), Formal Analysis (supporting), Investigation (supporting), Methodology (supporting), Software (supporting), Validation (supporting), Writing-original draft (supporting). **I.L. Quibod:** Formal Analysis (supporting), Methodology (supporting), Validation (supporting). **A.E.S. Almocera:** Formal Analysis (supporting), Validation (supporting). **L.A.E. Murao:** Formal Analysis (supporting), Supervision (supporting), Validation (supporting). **M.A.E Mata:** Conceptualization (lead), Formal Analysis (supporting), Funding Acquisition (lead), Methodology (supporting), Project Administration (lead), Resources (supporting), Supervision (lead), Validation (equal), Writing-review and editing (supporting). **R.C.H. del Rosario:** Conceptualization (supporting), Formal Analysis (supporting), Methodology (supporting), Resources (lead), Supervision (equal), Validation (equal), Writing-review and editing (lead).

REFERENCES

Aksamentov I, Roemer C, Hodcroft E, Neher R. Nextclade: clade assignment, mutation calling and quality control for viral genomes. J Open Source Softw 2021; 6(67):3773.

Amit AML, Pepito VCF, Dayrit MM. Early response to COVID-19 in the Philippines. West Pac Surveill Response J 2021; 12(1):56–60.

Amul GG, Ang M, Kraybill D, Ong SE, Yoong J. Responses to COVID-19 in Southeast Asia: diverse paths and ongoing challenges. Asian Econ Policy Rev 2022; 17(1):90–110.

Añonuevo LE, Lachica ZPT, Amistas DA, Lato JIE, Bontilao HLC, Catalan JMG, Pasion RJF, Yumang AP, Almocera AES, Arcede JP, Mata MAE, de los Reyes VAA. Transmission dynamics and baseline epidemiological parameter estimates of coronavirus disease 2019 pre-vaccination: Davao City, Philippines. PLoS One 2023; 18(3):e0283068.

Arcede JP, Caga-anan RL, Mammeri Y, Namoco RA, Gonzales ICA, Lachica ZP, Mata MAE. A modeling strategy for novel pandemics using monitoring data: the case of early COVID-19

pandemic in Northern Mindanao, Philippines. SciEnggJ 2022; 15:1–12.

Birnbaum ZW, Tingey FH. One-sided confidence contours for probability distribution functions. Ann Math Stat 1951; 22(4):592–596.

Bong CL, Brasher C, Chikumba E, McDougall R, Mellin-Olsen J, Enright A. The COVID-19 pandemic: effects on low- and middle-income countries. Anesth Analg 2020; 131(1):86–92.

Borda IJC, Buenaventura SGC, Lachica ZPT, Panogalinog IGM, Quibod IL, Almocera AES, Murao LAE, Mata MAE, del Rosario RCH. ph-ncr-phylodynamics-sarscov-2-v2. GitHub repository. 2024. Available from: https://github.com/amdabidss-health/ph-ncr-phylodynamics-sarscov-2-v2

Bouckaert R, Heled J, Kühnert D, Vaughan T, Wu CH, Xie D, Suchard MA, Rambaut A, Drummond AJ. BEAST 2: a software platform for Bayesian evolutionary analysis. PLoS Comput Biol 2014; 10(4):e1003537.

Danesh G, Elie B, Michalakis Y, Sofonea MT, Bal A, Behillil S, Destras G, Boutolleau D, Burrel S, Marcelin AG, Plantier JC, Thibault V, Simon-Loriere E, van der Werf S, Lina B, Josset L, Enouf V, Alizon S. Early phylodynamics analysis of the COVID-19 epidemic in France. Peer Community J 2021; 1:e45.

Department of Health. DOH confirms local transmission of COVID-19 in PH; reports 6th case. Manila: Department of Health, 2020. Available from: https://doh.gov.ph/press-release/doh-confirms-local-transmission-of-covid-19-in-ph-reports-6th-case/

Department of Health. COVID-19 case tracker Philippines. Manila: Department of Health, 2021. Available from: https://doh.gov.ph/diseases/covid-19/covid-19-case-tracker/

Drummond AJ, Bouckaert RR. Bayesian evolutionary analysis with BEAST. 1st ed. Cambridge: Cambridge University Press, 2015.

Drummond AJ, Rambaut A, Shapiro B, Pybus OG. Bayesian coalescent inference of past population dynamics from molecular sequences. Mol Biol Evol 2005; 22(5):1185–1192.

du Plessis L. SkylineTools TreeSlicer. GitHub repository. 2018. Available from: https://github.com/laduplessis/skylinetools

du Plessis L, McCrone JT, Zarebski AE, Hill V, Ruis C, Gutierrez B, Raghwani J, Ashworth J, Colquhoun R, Connor TR, Faria NR, Jackson B, Loman NJ, O'Toole Á, Nicholls SM, Parag KV, Scher E, Vasylyeva TI, Volz EM, Watts A, Bogoch II, Khan K, COVID-19 Genomics UK (COG-UK) Consortium, Aanensen DM, Kraemer MUG, Rambaut A, Pybus OG. Establishment and lineage dynamics of the SARS-CoV-2 epidemic in the UK. Science 2021; 371(6530):708–712.

Edrada EM, Lopez EB, Villarama JB, Salva Villarama EP, Dagoc BF, Smith C, Sayo AR, Verona JA, Trifalgar-Arches J, Lazaro J, Balinas EGM, Telan EFO, Roy L, Galon M, Florida CHN, Ukawa T, Villanueva AMG, Saito N, Nepomuceno JR, Ariyoshi K, Carlos C, Nicolasora AD, Solante RM. First COVID-19 infections in the Philippines: a case report. Trop Med Health 2020; 48:21.

Esguerra DJ. DOH: lack of test kits behind 'underreporting' of COVID-19 cases. INQUIRER.Net 2020. Available from: https://newsinfo.inquirer.net/1239007/doh-admits-lack-of-test-kits-behind-unintentional-underreporting-of-coronavirus-cases

Frost SDW, Volz EM. Viral phylodynamics and the search for an 'effective number of infections'. Philos Trans R Soc Lond B Biol Sci 2010; 365(1548):1879–1890.

Frost SDW, Pybus OG, Gog JR, Viboud C, Bonhoeffer S, Bedford T. Eight challenges in phylodynamic inference. Epidemics 2015; 10:88-92

Gillespie, DT. Exact stochastic simulation of coupled chemical reactions. J Phys Chem 1977; 81 (25): 2340–2361.

Giovanetti M, Cella E, Benedetti F, Rife Magalis B, Fonseca V, Fabris S, Campisi G, Ciccozzi A, Angeletti S, Borsetti A, Tambone V, Sagnelli C, Pascarella S, Riva A, Ceccarelli G, Marcello A, Azarian T, Wilkinson E, De Oliveira T, Alcantara LCJ, Cauda R, Caruso A, Dean NE, Browne C, Lourenco J, Salemi M, Zella D, Ciccozzi M. SARS-CoV-2 shifting transmission dynamics and hidden reservoirs potentially limit efficacy of public health interventions in Italy. Commun Biol 2021; 4:489.

Grenfell BT, Pybus OG, Gog JR, Wood JLN, Daly JM, Mumford JA, Holmes EC. Unifying the epidemiological and evolutionary dynamics of pathogens. Science 2004; 303(5656):327–332.

Hadfield J, Megill C, Bell SM, Huddleston J, Potter B, Callender C, Sagulenko P, Bedford T, Neher RA. Nextstrain: real-time tracking of pathogen evolution. Bioinformatics 2018; 34(23):4121–4123.

Hasell J, Mathieu E, Beltekian D, Macdonald B, Giattino C, Ortiz-Ospina E, Roser M, Ritchie H. A cross-country database of COVID-19 testing. Sci Data 2020; 7:345.

Haw NJL, Uy J, Sy KTL, Abrigo MRM. Epidemiological profile and transmission dynamics of COVID-19 in the Philippines. Epidemiol Infect 2020; 148:e204.

Inter-Agency Task Force. Inter-Agency Task Force resolution no. 56. Manila: Inter-Agency Task Force for the Management of Emerging Infectious Diseases, 2020a. Available from: https://boi.gov.ph/wp-content/uploads/2020/07/IATF-Resolution-No.-56_July-16-2020.pdf

Inter-Agency Task Force for the Management of Emerging Infectious Diseases. Inter-Agency Task Force resolution no. 37. Manila: Inter-Agency Task Force for the Management of Emerging Infectious Diseases, 2020b. Available from: https://www.pmapcebu.com/wp-content/uploads/2020/05/051520-IATF-Reso-37.pdf

Kanaka KK, Sukhija N, Goli RC, Singh S, Ganguly I, Dixit SP, Dash A, Malik AA. On the concepts and measures of diversity in the genomics era. Curr Plant Biol. 2023; 33:100278.

King AA, Nguyen D, Ionides EL. Statistical inference for partially observed Markov processes via the R package pomp. J Stat Softw 2016; 69(12):1–43.

Kühnert D, Stadler T, Vaughan TG, Drummond AJ. Simultaneous reconstruction of evolutionary history and epidemiological dynamics from viral sequences with the birth-death SIR model. J R Soc Interface 2014; 11(94):20131106.

Lam A, Duchene S. The Impacts of Low Diversity Sequence Data on Phylodynamic Inference during an Emerging Epidemic. Viruses 2021; 13(1):79.

Lau H, Khosrawipour T, Kocbach P, Ichii H, Bania J, Khosrawipour V. Evaluating the massive underreporting and

undertesting of COVID-19 cases in multiple global epicenters. Pulmonology 2021; 27(2):110–115.

Lim JS, Cho SI, Ryu S, Pak SI. Interpretation of the basic and effective reproduction number. J Prev Med Public Health 2020; 53(6):405-408.

Logrosa G, Mata MA, Lachica ZP, Estaña LM, Hassall M. Integrating risk assessment and decision-making methods in analyzing the dynamics of COVID-19 epidemics in Davao City, Mindanao Island, Philippines. Risk Anal 2022; 42(1):105–125.

Lu J, du Plessis L, Liu Z, Hill V, Kang M, Lin H, Sun J, François S, Kraemer MUG, Faria NR, McCrone JT, Peng J, Xiong Q, Yuan R, Zeng L, Zhou P, Liang C, Yi L, Liu J, Xiao J, Hu J, Liu T, Ma W, Li W, Su J, Zheng H, Peng B, Fang S, Su W, Li K, Sun R, Bai R, Tang X, Liang M, Quick J, Song T, Rambaut A, Loman N, Raghwani J, Pybus OG, Ke C. Genomic epidemiology of SARS-CoV-2 in Guangdong Province, China. Cell 2020; 181(5):997–1003.e9.

Marsaglia G, Tsang WW, Wang J. Evaluating Kolmogorov's distribution. J Stat Softw 2003; 8(18):1–4.

MathWorks Inc. MATLAB, version 9.11.0.1751886 (R2021b). Natick (MA): MathWorks Inc., 2021.

Minh BQ, Nguyen MAT, von Haeseler A. Ultrafast approximation for phylogenetic bootstrap. Mol Biol Evol 2013; 30(5):1188–1195.

Müller NF, du Plessis L. Tutorial using BEAST v2.7.x skyline plots. 2024. Available from: https://taming-the-beast.org/tutorials/Skyline-plots/Skyline-plots.pdf

Nadeau SA, Vaughan TG, Scire J, Huisman JS, Stadler T. The origin and early spread of SARS-CoV-2 in Europe. Proc Natl Acad Sci U S A 2021; 118(9):e2012008118.

Pajaron MC, Vasquez GNA. How effective is community quarantine in the Philippines? A quasi-experimental analysis. 2021. Available from: https://www.econstor.eu/bitstream/10419/230315/1/GLO-DP-0782.pdf

Parrish CR, Holmes EC, Morens DM, Park EC, Burke DS, Calisher CH, Laughlin CA, Saif LJ, Daszak P. Cross-species virus transmission and the emergence of new epidemic diseases. Microbiol Mol Biol Rev 2008; 72(3):457–470.

Philippine Statistics Authority. Highlights of the National Capital Region (NCR) population 2020 census of population and housing (2020 CPH). Quezon City: Philippine Statistics Authority, 2021. Available from: https://psa.gov.ph/content/highlights-national-capital-region-ncr-population-2020-census-population-and-housing-2020

Philippine Statistics Authority. Projected population, by age group, sex, and by single-calendar year interval, Philippines: 2010–2020 (medium assumption). Manila: Philippine Statistics Authority. Available from: https://psa.gov.ph/sites/default/files/attachments/hsd/pressrelease/Table4_9.pdf

Rambaut A. FigTree v1.4.4. Edinburgh: Institute of Evolutionary Biology, University of Edinburgh, 2018. Available from: http://tree.bio.ed.ac.uk/software/figtree/

Rambaut A, Lam TT, Carvalho LM, Pybus OG. Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). Virus Evol 2016; 2(1):vew007.

Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA. Posterior summarization in Bayesian phylogenetics using Tracer 1.7. Syst Biol 2018; 67(5):901–904.

Rhee C, Kanjilal S, Baker M, Klompas M. Duration of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) infectivity: when is it safe to discontinue isolation? Clin Infect Dis 2021; 72(8):1467–1474.

Research Institute for Tropical Medicine [RITM]. 2020-2021 RITM COVID-19 achievement report. 2021. Available from: https://ritm.gov.ph/wp-content/uploads/2023/02/2020-2021-COVID-19-Achievement-Report_FINAL.pdf

Rojas-Gallardo DM, Garzón-Castaño SC, Millán N, Jiménez-Posada EV, Martínez-Gutiérrez M, Ruiz-Saenz J, Cardona-Ospina JA. COVID-19 in Latin America: contrasting phylodynamic inference with epidemiological surveillance (molecular epidemiology of COVID-19 in Latin America). medRxiv 2020.

Sayers EW, Beck J, Bolton EE, Bourexis D, Brister JR, Canese K, Comeau DC, Funk K, Kim S, Klimke W, Marchler-Bauer A, Landrum M, Lathrop S, Lu Z, Madden TL, O'Leary N, Phan L, Rangwala SH, Schneider VA, Skripchenko Y, Wang J, Ye J, Trawick BW, Pruitt KD, Sherry ST. Database resources of the National Center for Biotechnology Information. Nucleic Acids Res 2021; 49(D1):D10–D17.

Seemann T, Lane CR, Sherry NL, Duchene S, Gonçalves da Silva A, Caly L, Sait M, Ballard SA, Horan K, Schultz MB, Hoang T, Easton M, Dougall S, Stinear TP, Druce J, Catton M, Sutton B, van Diemen A, Alpren C, Williamson DA, Howden BP. Tracking the COVID-19 pandemic in Australia using genomics. Nat Commun 2020; 11:4376.

Shu Y, McCauley J. GISAID: global initiative on sharing all influenza data – from vision to reality. Euro Surveill 2017; 22(13):30494.

Stadler T, Kühnert D, Bonhoeffer S, Drummond AJ. Birth–death skyline plot reveals temporal changes of epidemic spread in HIV and hepatitis C virus (HCV). Proc Natl Acad Sci U S A 2013; 110(1):228–233.

Tablizo F, Lapid C, Maralit B, Yap JM, Destura R, Alejandria M, Petronio-Santos JA, Morado EK, Dizon JG, Llames JH, Araiza SM, Punayan K, Velasco JM, Mejia JA, Dollete M, Salamat S, Tan C, Gabriel KA, Fabilloren SR, Demot B, Genavia S, Nipales J, Sanchez A, Gaza H, Arevalo G, Pangilinan C, Acosta S, Salinas M, Schwem B, Dela Tonga A, Galarion MJ, Dungca NT, Geganzo S, Bascos NA, Cutiongco-de la Paz EM, Saloma C. Analysis of SARS-CoV-2 Genome sequences from the Philippines: Genetic surveillance and transmission dynamics. medRxiv 2020. Available from: https://www.medrxiv.org/content/10.1101/2020.08.22.20180034v1.full.pdf

UP COVID-19 Pandemic Response Team. EndCOV: COVID-19 tracking dashboard. Quezon City: University of the Philippines, 2020. Available from: http://endcov.up.edu.ph/

Velasco JM, Chinnawirotpisan P, Joonlasak K, Manasatienkij W, Huang A, Valderama MT, Diones PC, Leonardia S, Timbol ML, Navarro FC, Villa II V, Tabinas Jr H, Chua Jr D, Fernandez S, Jones A, Klungthong C. Coding-Complete Genome Sequences of 23 SARS-CoV-2 Samples from the Philippines. Microbiol Resourc Announc 2020; 9(43):e01031-20.

Volz EM, Koelle K, Bedford T. Viral phylodynamics. PLoS Comput Biol 2013; 9(3):e1002947.

World Health Organization. WHO Director-General's opening remarks at the media briefing on COVID-19 - 11 March 2020. Geneva: World Health Organization, 2020a. Available from: https://www.who.int/director-general/speeches

World Health Organization. Coronavirus disease 2019 (COVID-19) situation report - 193. Geneva: World Health Organization, 2020b. Available from: https://www.who.int/publications/m/item/situation-report---193

World Health Organization. COVID-19 in the Philippines situation report 46. Manila: World Health Organization, 2020c. Available from: https://www.who.int/philippines/internal-publications-detail/covid-19-in-the-philippines-situation-report-46

World Health Organization. COVID-19 in the Philippines situation report 52. Manila: World Health Organization, 2020d. Available from: https://www.who.int/philippines/internal-publications-detail/covid-19-in-the-philippines-situation-report-52

World Health Organization. COVID-19 in the Philippines situation report 40. Manila: World Health Organization, 2020e. Available from: https://www.who.int/philippines/internal-publications-detail/covid-19-in-the-philippines-situation-report-40

Worobey M, Pekar J, Larsen BB, Nelson MI, Hill V, Joy JB, Rambaut A, Suchard MA, Wertheim JO, Lemey P. The emergence of SARS-CoV-2 in Europe and North America. Science 2020; 370(6516):564–570.

Yap PAG, Lachica ZP, Paras AL, Panogalinog IG, Tubay JM, Mata MA. An agent-based model of COVID-19 dynamics during enhanced community quarantine: exploring the role of food relief system in the presence of two SARS-CoV-2 variants. Front Appl Math Stat 2023; 9:1068180.

Zhu M, Shen J, Zeng Q, Tan JW, Kleepbua J, Chew I, Law JX, Chew SP, Tangathajinda A, Latthitham N, Li L. Molecular phylogenesis and spatiotemporal spread of SARS-CoV-2 in Southeast Asia. Front Public Health 2021; 9:685315.

Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, Zhao X, Huang B, Shi W, Lu R, Niu P, Zhan F, Ma X, Wang D, Xu W, Wu G, Gao GF, Tan W. A novel coronavirus from patients with pneumonia in China, 2019. N Engl J Med 2020; 382(8):727–733.