Mathematical Modelling of the Growth of SARS-CoV-2 (COVID-19) and SARS-CoV (SARS) Viruses in Vero E6 Cells

Mohd Yunus Abd Shukor1* and Mohammad Shafiul Alam2

1Department of Biochemistry, Faculty of Biotechnology and Biomolecular Sciences, Universiti Putra Malaysia, 43400 UPM Serdang, Selangor, Malaysia.
2Infectious Diseases Division, International Centre for Diarrhoeal Disease Research, Bangladesh (ICDDR,B), Dhaka-1212, Bangladesh.

*Corresponding author:
Prof. Dr. Mohd Yunus Shukor,
Department of Biochemistry,
Faculty of Biotechnology and Biomolecular Sciences,
Universiti Putra Malaysia,
43400 UPM Serdang, Selangor,
Malaysia.
Email: mohdyunus@upm.edu.my

INTRODUCTION

The SARS-CoV-2 virus is the agent causing the global pandemic COVID-19 which is still ravaging the human population with the current death tolls exceeding half a million people and more than 10 million people are infected [1]. Worldwide societies and economies have come to a near-complete standstill for the first time in a century because of this virus. The virus belongs to the family of coronaviruses (CoVs), which is known to have given rise to zoonotic infections over the past centuries on many occasions [2]. To date, seven CoVs segregated into two classes have been identified that can cause infections in humans. There are four endemic human CoVs (HCoVs) in one group, the first of which was reported in the 1960s, causing large numbers of common colds annually [3]. Three zoonotic CoVs form the other group that have recently caused outbreaks and include SARS— the severe acute respiratory syndrome coronavirus in 2002–2003, MERS— the Middle East respiratory syndrome-coronavirus since 2012 and COVID-19—the current SARS-CoV-2 pandemic. SARS-CoV-2 first emerged near the city of Wuhan in the People Republic of China in the fall of 2019 [2,4–6].

The newly emerging SARS-CoV-2 was quickly identified as a coronavirus with relatively close in genomic identity (~80 % identical) to the 2003 SARS-CoV virus. In addition, the organization of Open Reading Frames (ORFs) in the two viruses indicates no significant differences between these two parameters. However, the ymax or viral titer log (PFU/mL) upper asymptote values did not overlap suggesting significant differences between the two viruses with viral titer values for SARS-CoV (SARS) value higher than SARS-CoV-2 (COVID-19). The result obtained in this study warrants further study to the infection rate for both viruses that can be useful in studying potential similarity and differences between the two viruses.
have been extensively used in research for studying the cell culture-based infection models of SARS-CoV [12–14].

In a previous publication, the growth curves of SARS-CoV-2 and SARS-CoV in Vero E6 cells was compared for the first time. Both of the growth curves appears sigmoidal [12] and hence can be model using various standard primary models such as the popular modified Gompertz and modified Logistics [15–18]. The aim of this study is to apply the two models above to the growth curves of the viruses in Vero E6 cells.

MATERIALS AND METHODS

Data for Figure 1 were digitized using the software Webplotdigitizer 2.5 [19] from a published work [12]. Digitization using this software has been acknowledged for its reliability [20,21]. The data were then nonlinearly regressed using the curve-fitting software CurveExpert Professional software (v.1.6.3, http://www.curveexpert.net/, USA)) using two popular models often used to model microorganism growth— modified logistics (Equation 1) and modified Gompertz (Equation 2) [15,22].

\[
y = \frac{A}{1 + \exp \left[ \frac{4\mu m}{A}(\lambda-t)+2 \right]} \quad \text{(Eqn. 1)}
\]

\[
y = A\exp \left\{ -\exp \left[ \frac{\mu m e}{A}(\lambda-t)+1 \right] \right\} \quad \text{(Eqn. 2)}
\]

Note:
- \(A\) = Viral titer log (PFU/mL) lower asymptote;
- \(\mu\) = maximum specific viral titer log (PFU/mL) growth rate;
- \(\lambda\) = lag time (h.p.i);
- \(y_{max}\) = Viral titer log (PFU/mL) upper asymptote;
- \(e\) = exponent (2.718281828);
- \(t\) = sampling time (h.p.i)

Statistical analysis

Commonly used statistical discriminatory methods such as corrected AICc (Akaike Information Criterion), Root-Mean-Square Error (RMSE), and adjusted coefficient of determination (\(R^2\)) were utilized to find the best models from the two.

The RMSE was calculated according to Eqn. 1 [23], and a smaller number of parameters is expected to give a smaller RMSE value. \(n\) is the number of experimental data. \(Ob\) and \(Pd\) are the experimental and predicted data, while \(p\) is the number of parameters.

\[
RMSE = \sqrt{\frac{\sum_{i=1}^{n}(Pd_i - Ob_i)^2}{n - p}} \quad \text{(Eqn. 1)}
\]

As \(R^2\) or the coefficient of determination ignores the number of the parameter in a model, the adjusted \(R^2\) is utilized to overcome this issue. In the equation (Eqns. 2 and 3), the total variance of the \(y\)-variable is denoted by \(S_y^2\) while RMS is the Residual Mean Square.

\[
Adjusted(R^2) = 1 - \frac{RMSE}{S_y^2} \quad \text{(Eqn. 2)}
\]

\[
Adjusted(R^2) = 1 - \frac{(1 - R^2)(n-1)}{(n-p-1)} \quad \text{(Eqn. 3)}
\]

The Akaike Information Criterion (AIC) is based on information theory. It balances between the goodness of fit of a particular model and the complexity of a model [24]. To handle data having a high number of parameters or a smaller number of values corrected Akaike information criterion (AICc) is utilized [25]. The AICc is calculated as follows (Eqn. 4), where \(p\) signifies the quantity of parameters and \(n\) signify the quantity of data points. A model with a smaller value of AICc is deemed likely more correct [25].

\[
AICc=2p+\ln\left(\frac{RSS}{n}\right)+2\left(\frac{p+1}{n-p-2}\right) \quad \text{(Eqn. 4)}
\]

RESULTS AND DISCUSSIONS

Of the two models, the best performance was the Logistic model with the lowest value for RMSE, AICc and the highest value for adjusted \(R^2\) (Table 1). The Logistic model was then utilized to model the growth data (Fig. 1). The result indicates maximal growth occurred after 15 h.p.i and the specific growth rate appears slightly higher for SARS-CoV than SARS-CoV-2 but this require analysing the confidence interval value. The coefficients for the model are shown in Table 2.

Table 1. Error function analysis between the two models utilized to model the growth curves of SARS-CoV and SARS-CoV-2 in Vero E6 cells.

<table>
<thead>
<tr>
<th>Model</th>
<th>SARS-CoV RMSE</th>
<th>SARS-CoV-2 RMSE</th>
<th>SARS-CoV AICc</th>
<th>SARS-CoV-2 AICc</th>
</tr>
</thead>
<tbody>
<tr>
<td>Modified</td>
<td>0.9533</td>
<td>0.5227</td>
<td>-33.08</td>
<td>0.8921</td>
</tr>
<tr>
<td>Gompertz</td>
<td>0.9069</td>
<td>0.4247</td>
<td>-46.34</td>
<td>0.9194</td>
</tr>
<tr>
<td>Modified</td>
<td>0.9692</td>
<td>0.4247</td>
<td>-46.34</td>
<td>0.9194</td>
</tr>
</tbody>
</table>

Fig. 1. Modelling of the growth curves of SARS-CoV and SARS-CoV-2 in Vero E6 cells. Solid black line indicates best fit curve whilst solid red and blue lines represent 95% confidence interval bands for SARS-CoV and SARS-CoV-2, respectively.
The overlapping of the 95% confidence intervals for the parameters \( \lambda \) (lower asymptote value of viral titer) and \( \mu_0 \) (maximum specific viral titer) in the modified Logistics model (Table 2) which can also be visualized (Fig. 1) indicate no significant differences [26] between the parameters. However, the \( y_{\text{max}} \), or viral titer log (PFU/mL) upper asymptote values did not overlap suggesting significant differences between the two values with viral titer values for SARS-CoV (SARS) value higher than SARS-CoV-2 (COVID-19). The biological significance of this finding needs further studies especially when a mathematical model of infection study in human infection has shown that the growth rate of SARS-CoV-2 (COVID-19) is double to that of SARS-CoV (SARS) [27]. Both of the logistic and the Gompertz curves have found utility in modelling the growth of HIV-1 positive population in Finland in population dynamics [28] and in in vitro growth of several viruses [29–31].

### CONCLUSION

In this work, the modelling of the growth of the viruses SARS-CoV-2 and SARS-CoV in Vero E6 cells is reported for the first time. Two commonly used models; modified Gompertz and modified Logistics were utilized and of the two, the best performance was the logistic model with the lowest value for RMSE, AICc and the highest value for adjusted \( R^2 \). The logistic models revealed overlapping of the 95% confidence intervals for the parameters \( \lambda \) (lower asymptote value of viral titer) and \( \mu_0 \) (maximum specific viral titer) indicating no significant differences between the parameters. However, the \( y_{\text{max}} \), or viral titer log (PFU/mL) upper asymptote values did not overlap suggesting significant differences between the two values with viral titer values for SARS-CoV (SARS) value higher than SARS-CoV-2 (COVID-19). Further refinement to the experiment can be done by introducing bootstrapping and Monte Carlo simulation on the overlapped confidence interval values of the two parameters to assess significance.

### REFERENCES


5. Shukor MS, Shukor MY. Statistical diagnostic tests of residuals from the Gompertz model used in the fitting of the growth of E. coli measured using a real-time impedimetric biosensor. Nanobio E. coli from the Gompertz model used in the fitting of the growth of Acinetobacter sp. strain DRY12 grown on diesel. J Environ Bioremediation Toxicol. 2014;2(1):33–7.


7. 此文引用文献略。


