

Non-equilibrium critical dynamics of RNA virus evolution*

Xiaofei Ge,[†] Kaichao You,[‡] Zeren Tan,[§] Hedong Hou,[¶] Yang Tian,^{**} and Pei Sun^{††}

RNA virus evolves in a complex manner. Here we present a theory to characterize the RNA virus evolution as a physical system with absorbing states and avalanche behaviors, trying to bridge between biological phenomenons (e.g., phylogenetic tree and infection) and physics concepts (e.g., directed percolation and self-organized criticality). The global epidemic of COVID-19 offers a unique opportunity as a natural experiment to verify our theory. We find that SARS-CoV-2 exhibits scale-invariant avalanches as mean-field theory predictions. The observed scaling relation, universal collapse, and slowly decaying auto-correlation suggest a non-equilibrium critical dynamics of SARS-CoV-2 evolution. Interestingly, the lineages that emerge from critical evolution processes coincidentally match with threatening lineages of SARS-CoV-2 (e.g., the Delta virus). Although being validated only on SARS-CoV-2 yet, the present theory may be a general formalism to portray the evolution of RNA virus.

With high mutation and replication capacities, RNA viruses exhibit intricate evolution [1–4]. Studying the elusive evolution dynamics of RNA virus remains as an attractive challenge in biology, as RNA virus evolution is a frequent cause of new pathogens and may be a significant route in medicines [5–7]. Meanwhile, RNA virus interests physicians for its potential to serve as a representative model of long-term molecular evolution [8, 9].

Till now, the physics underlying RNA virus evolution remains largely unknown. Although mutations and natural selection pressures (e.g., immune) emerge in seconds, small changes in RNA virus need long-term accumulations to be evident [10]. Therefore, an analytic theory of RNA virus evolution is required to non-trivially handle multiple time scales with extreme differences in magnitude orders. While important progress has been accomplished in developing phenomenological models to reduce model complexity [9, 11–24], these works are powerful in imitating phenomena but limited in explaining mechanisms. Interpretations of phenomenological parameters with experiments should be prudently considered. Notably, several studies have tried to compromise between analytic theories and phenomenological models [10].

From the physical perspective, previous works usually aim at elaborating RNA virus evolution in the context of population dynamics. For instance, the Nowak–May HIV model considers the life cycles (e.g., death) of cells and free HIVs while analyzing infections [15]. The terminology “evolution” mainly refers to the dynamics of mutant virus population size [10, 17] and is applicable when detailed virus amount is of interest. However, there is another non-negligible aspect of evolution remaining for exploration. From the biological perspective, the dynamics accountable for phylogenetic tree formation is informative for the global evolution properties [25–28]. Compared with the evolution considered previously, this unexplored aspect may convey key information about the macroscopic laws of RNA virus evolution.

The objectives of the current research are twofold. We aim at exploring the global properties of RNA virus evolution by developing a physics theory of the dynamics

underlying phylogenetic tree formation. The theory is demanded to map fundamental biological factors to general characterizations of system dynamics so as to guarantee its applicability on real data and universality to arbitrary RNA viruses. Applying this theory, we present a systematic analysis of SARS-CoV-2, the virus accounting for COVID-19 [29]. The global infection of SARS-CoV-2 provides us a unique opportunity with rich data (e.g., a sufficiently large phylogenetic tree) to validate our theory. Moreover, confirming the macroscopic characteristics of SARS-CoV-2 evolution may benefit to developing coping strategies of COVID-19.

Backgrounds of RNA virus.—With reasonable errors, the mutation rate σ of single RNA virus is typically

$$\sigma \in \left[10^{-5} \chi \phi \frac{1}{\tau}, 10^{-4} \chi \phi \frac{1}{\tau} \right], \quad (1)$$

where $\chi \in \mathbb{N}^+$ measures the number of nucleotides, parameter τ denotes entrance and eclipse periods, and $\phi \simeq 10^3$ denotes burst size (the number of viruses produced by a host cell) [1, 9, 30]. These parameters are principally constant for a given RNA virus. Note that the original RNA virus vanishes after replication [31].

We then turn to analyze RNA virus populations. The replication implies the vanishing of the original RNA virus and the birth of new RNA viruses that can also replicate themselves. Under ideal conditions, the cumulative mutation rate of a RNA virus population of initial size $\omega \in \mathbb{N}^+$ during a duration of $[0, \kappa]$ (assume τ is divisible by κ) has a supremum

$$\sup_{\omega} \zeta(\omega) = \omega \sigma \int_0^{\kappa} [\phi^{\lfloor t/\tau \rfloor} - \phi^{\lfloor t/\tau \rfloor - 1}] dt, \quad (2)$$

$$= \omega \sigma \tau \left(\phi^{\lfloor \kappa/\tau \rfloor - 1} - \frac{1}{\phi} \right), \quad (3)$$

where $\lfloor \cdot \rfloor$ denotes the floor function. The supremum is reached when each RNA virus replicates and mutates independently and freely (e.g., with sufficient susceptible cells). However, these ideal conditions are not realistic since host cell resources may be limited. Therefore, bio-

logically sound models of RNA virus population dynamics, including the discussed previous works [9–24], bound the proliferation of RNA virus by time-dependent host cell resource $\lambda(t)$ (e.g., the number of susceptible cells)

$$\sup_{\omega} \widehat{\zeta}(\omega) = \omega \sigma \int_0^{\kappa} \min \left[\lambda(t), \phi^{\lfloor \frac{t}{\tau} \rfloor} - \phi^{\lfloor \frac{t}{\tau} \rfloor - 1} \right] dt. \quad (4)$$

Although Eq. (4) can not be solved until $\lambda(t)$ is given, we can know

$$\sup_{\lambda} \sup_{\omega} \widehat{\zeta}(\omega) = \sup_{\omega} \zeta(\omega). \quad (5)$$

The difference $\sup_{\omega} \zeta(\omega) - \sup_{\omega} \widehat{\zeta}(\omega)$ will be minimized when $\lambda(t) \geq \phi^{\lfloor t/\tau \rfloor} - \phi^{\lfloor t/\tau \rfloor - 1}$.

In summary, host cell resource $\lambda(t)$ intrinsically determines the cumulative daily mutation rate of a RNA virus population. Instead of proposing detailed definitions of $\lambda(t)$ as previous studies (e.g., define life circles of cells) [9–24], we leave the characterization of $\lambda(t)$ as a task of epidemic and observe the evolution dynamics of RNA virus controlled by $\lambda(t)$ directly from real data.

Theory of RNA virus evolution.—Certainly, one can not expect to count $\lambda(t)$ in a real epidemic. This parameter can be reasonably replaced by the number of human hosts. Assuming that a lineage l_i of RNA virus emerges at moment t' and one of its progeny lineage l_j is born at moment t'' , we can measure the duration length of evolution as $T = t'' - t'$. Meanwhile, we define $S(T)$ as the number of cumulative confirmed patients with lineage l_i during $[t', t'']$ and denote the time-dependent number of confirmed patients with lineage l_i by $S(t | T)$ (here $t \in [t', t'']$). The terminology “lineage” can be generally understood as a kind of definition of virus sub-type. Viruses of lineage l_i need to accumulate sufficient mutations to make lineage l_j emerge.

We can relate these concepts with a branching process concerning the number of copies of l_i during $[t', t'']$, that is, a process where a patient with lineage l_i infects multiple people before the virus mutates to lineage l_j during replication. The realisation of this branching process, from the initialization t' to the termination t'' after which a new lineage l_j emerges, is referred to as an avalanche. The size and life time of avalanche are exactly $S(T)$ and T . This branching process is slightly different from the standard one, where the termination is defined as a moment after which the population size (e.g., the number of patients with lineage l_i) remains 0 indefinitely [32–34].

We can further relate the branching process of evolution with directed percolation, a universality class exhibiting a phase transition separating an absorbing state from an active state [32, 35, 36]. In directed percolation (e.g., the contact process [37]), sites (e.g., people) can be either active (infected) or inactive (healthy). With different balance between infection and recovery, the infection process may propagate over the system or vanish gradually. If infection vanishes, the system is trapped in a

completely inactive state, the so-called absorbing state. Directed percolation is a non-equilibrium process since detailed balance breaks in the absorbing state (it can be reached but not be escaped) [38].

We are interested in potential self-organized criticality (SOC) due to its pervasiveness in biological systems (e.g., neural avalanches in the brain [39–41]). The self-organization to criticality distinguishes SOC (e.g., Refs. [42, 43]) from ordinary critical phenomena [38, 44]. At criticality, the system jumps between absorbing configurations by avalanches [38, 44]. From the biological perspective, this property implies an unpredictable and unpreventable evolution of RNA virus. Numerous mean field theories of directed percolation have been proposed to determine avalanche exponents to study the scale-invariance underlying P_T and P_S , the probability distributions of T and S [33, 45–55]. Note that S is an abbreviation of $S(T)$. Here we re-calculate avalanche exponents following a similar idea of Refs. [32, 49, 56].

Consider a time-continuous infection process where the first patient emerges at t' . A random patient at moment $t \in [t', t'']$ implies three possibilities: becoming absorbed with probability ρ , creating another patient with probability θ , or remaining effect-free with probability $1 - (\rho + \theta)$. In critical states, we have $\rho = \theta$ [49]. We define $\mathcal{A}_n(t)$ as the probability for n patients to exist at $t^* + t$ given that 1 patient exists at t^* . Assuming the independence of patient emergence, we have

$$\mathcal{A}_n(t) = \sum_{n_1 + \dots + n_k = n} \mathcal{A}_{n_1}(t) \dots \mathcal{A}_{n_k}(t). \quad (6)$$

Assume $\mathcal{A}_n(t), n \in \mathbb{N}^+$ admit a Maclaurin expansion $\mathcal{A}_k(t) = a_k t + o(t^2)$ (when $k \neq 1$) or $\mathcal{A}_k(t) = a_k t + 1 + o(t^2)$ (when $k = 1$) where $a_k = d\mathcal{A}_n(0)/dt$, we have $a_0 = a_2 = \rho$ and $a_1 = -2\rho$ [49]. Meanwhile, we can know

$$\mathcal{A}_n(t + dt) - \mathcal{A}_n(t) = \sum_{k=0}^{\infty} a_k \mathcal{A}_{n-k}(t) dt. \quad (7)$$

Eqs (6-7) readily lead to

$$\begin{aligned} \frac{\partial}{\partial t} \mathcal{G}(s, t) &= \sum_{k=0}^{\infty} a_k \sum_{n=0}^{\infty} \left(\sum_{n_1 + \dots + n_k = n-k} \prod_{i=1}^k \mathcal{A}_{n_i}(t) \right) s^n, \\ \end{aligned} \quad (8)$$

$$= \sum_{k=0}^{\infty} a_k \mathcal{G}(s, t)^k, \quad (9)$$

where $\mathcal{G}(s, t) = \sum_{n=0}^{\infty} \mathcal{A}_n(t) s^n$, $s \in [0, 1]$ denotes the generating function. Applying a trick introduced in Ref. [49], we define

$$\mathcal{H}(s) = \frac{\partial}{\partial t} \mathcal{G}(s, 0), \quad (10)$$

which naturally leads to

$$\frac{\partial}{\partial t} \mathcal{G}(s, t) = \mathcal{H}(\mathcal{G}(s, t)), \quad (11)$$

the backward Chapman-Kolmogorov equation. Meanwhile, $\mathcal{H}(s) = \rho(1-s)^2$ can be derived based on a_0 , a_1 , and a_2 [49]. Taken together, we have

$$\frac{\partial}{\partial t} \mathcal{G}(s, t) = \rho(1 - \mathcal{G}(s, t))^2. \quad (12)$$

Note that the initial condition is $\mathcal{G}(s, 0) = s$ since the first patient emerges at t^* . Solving Eq. (12), we derive an analytic expression

$$\mathcal{G}(s, t) = \frac{\rho(1-s)(t-t^*)}{\rho(1-s)(t-t^*)+1}. \quad (13)$$

Therefore, we have $\mathcal{A}_0(t) = \mathcal{G}(0, t) = \frac{\rho(t-t^*)}{\rho(t-t^*)+1}$, supporting a calculation of lifetime distribution $\mathcal{P}_T(t)$

$$\lim_{t \rightarrow \infty} \mathcal{P}_T(t) = \lim_{t \rightarrow \infty} \frac{d}{dt} \mathcal{G}(0, t) \sim t^{-2}. \quad (14)$$

Following Ref. [32, 49, 56], one can similarly calculate

$$\lim_{s \rightarrow \infty} \mathcal{P}_S(s) \sim s^{-\frac{3}{2}}. \quad (15)$$

These exponents are consistent as the predictions of other mean field theories [33, 45–55].

To verify whether the system is at criticality, we need to consider a scaling relation [38, 44, 57, 58]

$$\gamma = \frac{\alpha - 1}{\beta - 1}, \quad (16)$$

where $\mathcal{P}_T(t) \propto t^{-\alpha}$, $\mathcal{P}_S(s) \propto s^{-\beta}$, and $\langle S \rangle(T) \propto T^\gamma$. Inserting $\alpha = 2$ and $\beta = \frac{3}{2}$ into Eq. (16), we derive $\gamma = 2$. Meanwhile, a more precise verification can be implemented by the collapse shape [39, 58, 59]

$$\langle \mathcal{V}(t | T) \rangle = \mathcal{F}\left(\frac{t}{T}\right), \quad (17)$$

in which the expectation is averaged across different T . Notion $\mathcal{F}(\cdot)$ denotes a universal scaling function. Notion $\mathcal{V}(t | T)$ is defined as the average collapse shape of all avalanches with the same life time T

$$\mathcal{V}(t | T) = T^{1-\gamma} \langle S(t | T) \rangle. \quad (18)$$

where $S(t | T)$ denotes the time-dependent avalanche size during these avalanches. At criticality, all data of $\mathcal{V}(t | T)$ should collapse onto $\mathcal{F}(\cdot)$, a parabolic function, with reasonable errors [39, 58, 59]. In some cases, function $\mathcal{F}(\cdot)$ may contain both the parabolic component and a slight global trend (e.g., increasing or decreasing), making the shape not perfectly parabolic. De-trending can be used to deal with this issue during data pre-processing. Moreover, a more practical verification concerns the slow and exponential decay of auto-correlation [41, 60]

$$\ln \left[\frac{\text{Cov}(S(t_i | T), S(t_j | T))}{\text{Cov}(S(t_i | T), S(t_i | T))} \right] = -\xi \left(\frac{t_j - t_i}{T} \right) + c, \quad (19)$$

where $t_i \in [t', t'']$ and $t_j \in [t_i, t'']$.

Evolution of SARS-CoV-2.—Since December 2019, COVID-19, a kind of contagious disease caused by SARS-CoV-2, has spread worldwide [29]. We use an open-source database [61] to acquire the necessary data. The criterion of lineage definition can be seen in [62]. Our data includes all lineages that emerge before 27 December 2021.

In the SARS-CoV-2 database [61], the probability for a patient to be infected with each lineage of virus is estimated with confidence intervals. By multiplying the midpoint of confidence interval with the total number of worldwide patients recorded by the World Health Organization (WHO) [63], we can estimate the accumulative and daily numbers of confirmed patients with each lineage of virus, corresponding to $S(T)$ and $S(t | T)$ of each lineage (see Fig. 1a). Similarly, we can estimate the upper and lower bounds of $S(T)$ and $S(t | T)$ based on the upper and lower bounds of confidence interval.

In Fig. 1b, we show the probability distributions of life time T and avalanche size S (for midpoint and bounds). These distributions are derived after a fine-grained data binning process with 1000 bins is applied, which is useful in de-noising while controlling information loss. We discover that $\mathcal{P}_T(\cdot)$ does not follow power-law properties for all values of T . In real cases, the power-law may only hold on specific tails $\{\mathcal{P}(X) | X \geq X'\}$ of the empirical distribution of variable X , where X' denotes a certain distribution cutoffs [64]. Therefore, distribution cutoff estimation is necessary for T and S . We apply the method introduced in Ref. [65] to estimate T' , S' , S'_- (lower bound), and S'_+ (upper bound), which calculates the Kolmogorov-Smirnov (KS) statistic η between the observed distribution above cutoff and a standard power-law distribution. An ideal cutoff is expected to minimize η [65]. To control the reduction of sample size, we relatively relax the restriction and estimate cutoffs when $\langle \eta \rangle - 0.5 \text{ std}(\eta)$ is reached at the first time, where $\text{std}(\cdot)$ denotes the standard deviation (Fig. 1c). Distribution cutoffs are estimated as $T' = 29$, $S' = 13320$, $S'_- = 12760$, and $S'_+ = 13980$. The distribution tails above cutoffs cover 93.879%, 76.38%, 70.526%, and 91.683% of original samples, supporting robust estimations of power-law exponents. A maximum likelihood estimation of power-law exponent [65] is implemented on the samples above cutoffs. An ideal exponent can maximize the likelihood \mathcal{L} (Fig. 1c) [65]. In our results, exponents are estimated as $\alpha = 1.69$ ($v = 7.13\%$), $\beta = 1.32$ ($v = -4.53\%$), $\beta_- = 1.32$ ($v = -5.06\%$), and $\beta_+ = 1.34$ ($v = -14.88\%$). To calculate the average KS statistic η^* between sample distributions and estimated models, we first generate 1000 sample distributions (each sample distribution contains $n \in [500, 5000]$ samples) of estimated power-law models. Then, we define $v = \frac{\hat{\eta} - \eta^*}{\eta^*}$ to reflect the goodness of estimation, where $\hat{\eta}$ is the KS statistic between estimated power-law models and empirical distributions

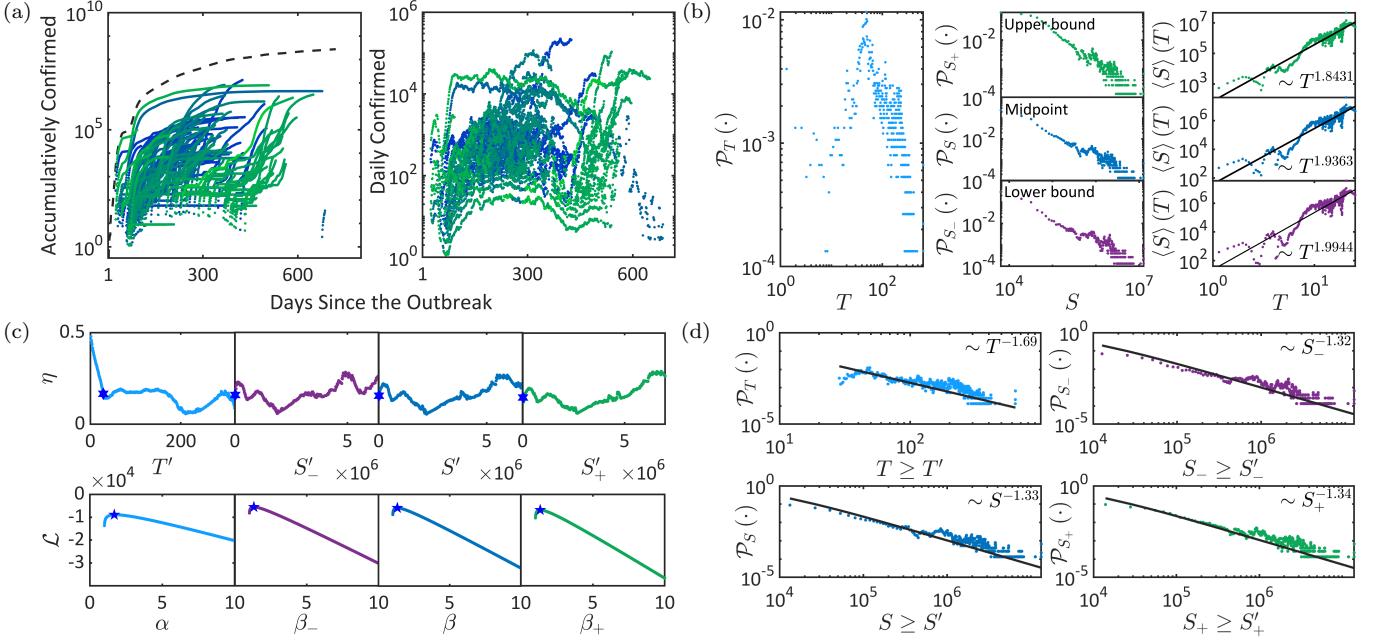


FIG. 1: Avalanches of SARS-CoV-2 evolution. (a) Numbers of accumulatively and daily confirmed patients with viruses of different lineages (scatters with different colors) and the total accumulative number of patients (black dashed line). (b) Data distributions of $\mathcal{P}_T(\cdot)$ vs. T , $\mathcal{P}_S(\cdot)$ vs. S , and $\langle S \rangle(T)$ vs. T . (c) The KS statistic η for estimating distribution cutoffs and the likelihood \mathcal{L} for estimating avalanche exponents. Blue stars denote estimated values. (d) The power-law distributions of T and S (midpoint and bounds of confidence interval) above cutoffs.

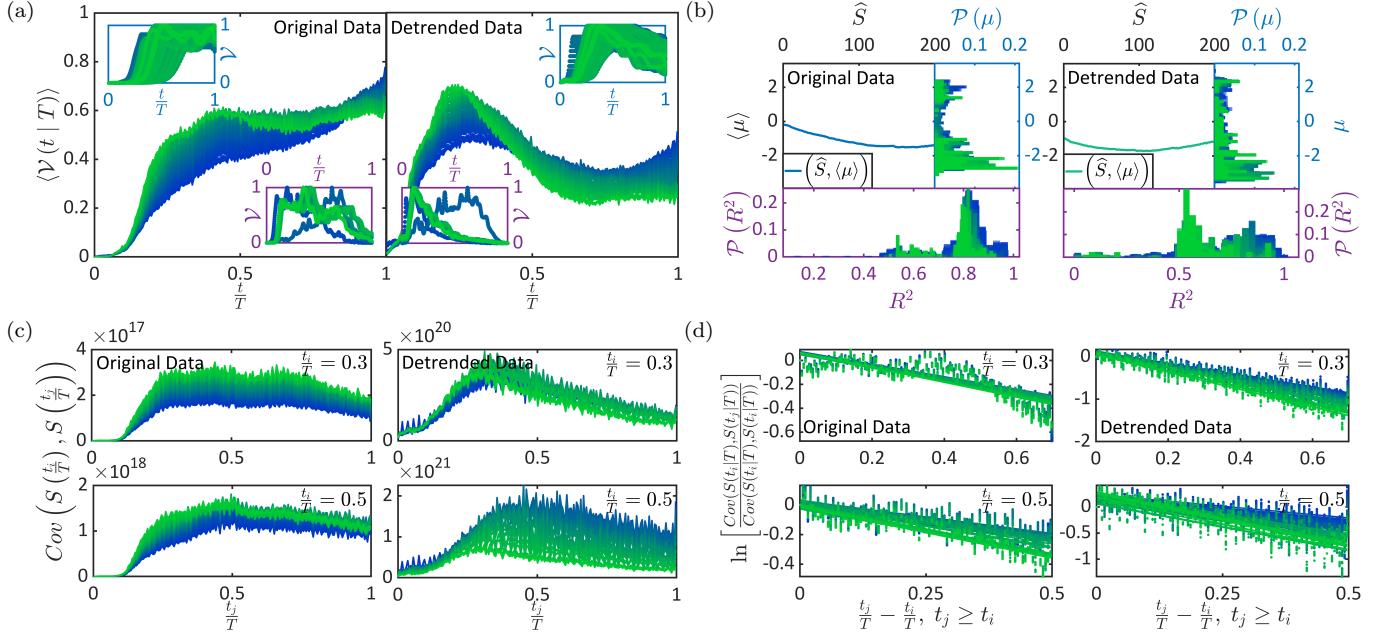


FIG. 2: Criticality of SARS-CoV-2 evolution. (a) The collapse shape $\langle \mathcal{V}(t | T) \rangle$ vs. $\frac{t}{T}$ of original and detrended data are shown (main plots), accompanied by several examples of $\mathcal{V}(t | T)$ vs. $\frac{t}{T}$ (inserted small plots, blue boxes correspond to avalanches with $T \in [50, 200]$ and $S \geq 50S'$ while purple boxes correspond to avalanches with $T \geq 400$ and $S \geq 50S'$). (b) The average quadratic coefficient $\langle \mu \rangle$ shown as a function of \widehat{S} (black boxes), the probability distribution of μ (blue boxes), and the R^2 statistics of curve fitting (purple boxes). (c-d) The auto-correlation functions of original and detrended data and their exponential decays.

above cutoffs. We suggest that $v < 10\%$ can be a rea-

sonable standard for ideal estimations. Our estimated

models are shown in Fig. 1d.

Meanwhile, we implement least square fitting on $\langle S \rangle(T)$ vs. T to derive $\gamma_+ = 1.8431$ ($R^2 = 0.9112$), $\gamma = 1.9363$ ($R^2 = 0.8871$), and $\gamma_- = 1.9944$ ($R^2 = 0.8652$) in Fig. 1b. Together with our estimated avalanche exponents α and β , these results coincide with the scaling relation in Eq. (8), suggesting potential criticality.

To analyze criticality more precisely, we calculate collapse shape following Eqs. (17-18). We further verify if the estimators of distribution cutoffs for T and S affects collapse shape since we have relaxed restrictions during estimation. In Fig. 2a, we show $\langle V(t|T) \rangle$ vs. $\frac{t}{T}$ of avalanches with $T \geq T'$ and $S \geq \hat{S}$, where $\hat{S} \in \{S', \dots, 200S'\}$ (the main panel of Fig. 2a). By increasing \hat{S} , we can simultaneously increase the distribution cutoff of T . Our results suggest that the trend of $\langle V(t|T) \rangle$ is principally independent of \hat{S} . Meanwhile, we notice that $\langle V(t|T) \rangle$ exhibits a mixture of global increasing trend and parabolic trend. The increasing trends may reflect the intrinsic property of the raw data. Therefore, we recalculate S by multiplying the midpoint of confidence interval [61] with the average total number of worldwide patients in the WHO database [63] for detrending. In Fig. 2a, $\langle V(t|T) \rangle$ plausibly exhibits a parabolic trend in the detrended data. Meanwhile, we show several instances of $V(t|T)$ vs. $\frac{t}{T}$ in Fig. 2a. In Fig. 2b, the parabolic trend of $\langle V(t|T) \rangle$ is verified quantitatively by quadratic polynomial fitting. We show the average quadratic coefficient $\langle \mu \rangle$ (averaged across avalanches under each condition of \hat{S}) as a function of \hat{S} . Consistent with Fig. 2a, $\langle \mu \rangle < 0$ holds for every \hat{S} in both original and detrended data, suggesting plausible parabolic trends of $\langle V(t|T) \rangle$. The probability distributions of $\langle \mu \rangle$ and the goodness of fitting, R^2 , are also shown.

In Fig. 2c, we show the auto-correlation (covariance) $\text{Cov}(S(t_i|T), S(t_j|T))$ under each condition of \hat{S} , where we select $t_i \in \{0.3T, 0.5T\}$ as instances. In Fig. 2d, the corresponding decays of auto-correlation are fitted to derive ξ in Eqs. (19). In the raw data, we obtain $\langle \xi \rangle = -0.5407$ (averaged $R^2 = 0.6301$) for $t_i = 0.3T$ and $\langle \xi \rangle = -0.5159$ (averaged $R^2 = 0.6324$) for $t_i = 0.5T$. In the detrended data, we find that $\langle \xi \rangle = -1.6679$ (averaged $R^2 = 0.6301$) for $t_i = 0.3T$ and $\langle \xi \rangle = -1.4251$ (averaged $R^2 = 0.6332$) for $t_i = 0.5T$. These results suggest exponential decays of auto-correlation with reasonable errors, offering a practical verification of potential criticality.

Interestingly, our results coincide with threatening lineages of SARS-CoV-2 if we filter directed edges in its original phylogenetic tree (edges correspond to avalanches and indicate evolution paths of lineages) as following: an edge is removed unless its avalanche satisfies $S \geq S'$ and $T \geq T'$ (Fig. 3). The filtered result, referred to as critical phylogenetic tree, is a sub-graph of the original one that includes only critical avalanches. Numerous lineages

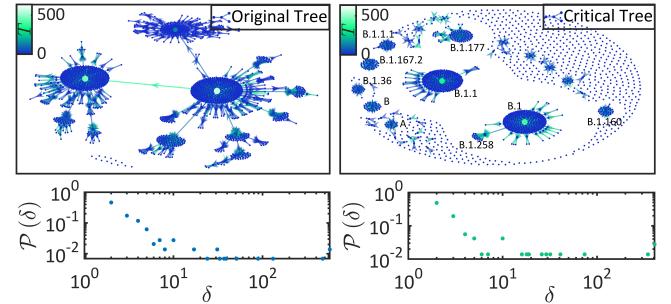


FIG. 3: Original and critical phylogenetic trees. Edges are colored according to T of corresponding avalanches.

with high out-degrees δ (number of offspring lineages) in the original tree become isolated. Threatening lineages, such as B.1.167.2 (the Delta virus), B.1.1 (the parent of Alpha, Beta, and Omicron viruses), and B.1 (the parent of Epsilon, Eta, Iota, and Mu viruses), can be found by searching lineages with high δ in the critical tree.

In this paper, we have presented a directed percolation theory to characterize RNA virus evolution as a system with absorbing states and avalanches. Applying the theory, we discover that SARS-CoV-2 evolution may proceed at criticality. Threatening lineages can be coincidentally found on the critical phylogenetic tree. These findings suggest the potential of our theory in analyzing RNA virus evolution and related phenomena.

Acknowledgement: This project is supported by the Artificial and General Intelligence Research Program of Guo Qiang Research Institute at Tsinghua University (2020GQG1017) as well as the Tsinghua University Initiative Scientific Research Program.

*Correspondence should be addressed to Y.T and P.S.

[†]gxlf2@mails.tsinghua.edu.cn; School of Pharmaceutical Sciences, Tsinghua University, Beijing, 100084, China.

[‡]ykc20@mails.tsinghua.edu.cn; School of Software, Tsinghua University, Beijing, 100084, China.

[§]tanzr2@mails.tsinghua.edu.cn; Institute for Interdisciplinary Information Science, Tsinghua University, Beijing, 100084, China.

[¶]hedong.hou@etu.u-paris.fr; UFR de Mathématiques, Université de Paris, 75013 Paris, France

^{**}tiany20@mails.tsinghua.edu.cn; Department of Psychology & Tsinghua Laboratory of Brain and Intelligence, Tsinghua University, Beijing, 100084, China.

^{††}peisun@tsinghua.edu.cn; Department of Psychology & Tsinghua Laboratory of Brain and Intelligence, Tsinghua University, Beijing, 100084, China.

[1]E. Domingo, C. Escarmís, N. Sevilla, A. Moya, S. F. Elena, J. Quer, I. S. Novella, and J. J. Holland, Basic concepts in rna virus evolution, *The FASEB Journal* **10**, 859 (1996).

[2]E. Domingo and J. Holland, Rna virus mutations and fitness for survival, *Annual review of microbiology* **51**, 151

- (1997).
- [3] P. T. Dolan, Z. J. Whitfield, and R. Andino, Mechanisms and concepts in rna virus population dynamics and evolution, *Annual Review of Virology* **5**, 69 (2018).
- [4] S. Duffy, Why are rna virus mutation rates so damn high?, *PLoS biology* **16**, e3000003 (2018).
- [5] S. Ojosnegros and N. Beerenwinkel, Models of rna virus evolution and their roles in vaccine design, *Immunome research* **6**, 1 (2010).
- [6] K. K. Irwin, N. Renzette, T. F. Kowalik, and J. D. Jensen, Antiviral drug resistance as an adaptive process, *Virus evolution* **2** (2016).
- [7] E. Jakobsson, A. B. Farimani, E. Tajkhorshid, and N. Aluru, Combining physics-based and evolution-based methods to design antibodies against an evolving virus, *Biophysical Journal* **118**, 482a (2020).
- [8] K. Sneppen and G. Zocchi, *Physics in molecular biology* (Cambridge University Press, 2005).
- [9] L. S. Tsimring, H. Levine, and D. A. Kessler, Rna virus evolution via a fitness-space model, *Physical review letters* **76**, 4440 (1996).
- [10] A. Korobeinikov, A. Archibasov, and V. Sobolev, Multi-scale problem in the model of rna virus evolution, in *Journal of Physics: Conference Series*, Vol. 727 (IOP Publishing, 2016) p. 012007.
- [11] W. Fontana, W. Schnabl, and P. Schuster, Physical aspects of evolutionary optimization and adaptation, *Physical Review A* **40**, 3301 (1989).
- [12] A. Sasaki, Evolution of antigen drift/switching: continuously evading pathogens, *Journal of Theoretical Biology* **168**, 291 (1994).
- [13] Y. Haraguchi and A. Sasaki, Evolutionary pattern of intra-host pathogen antigenic drift: effect of cross-reactivity in immune response, *Philosophical Transactions of the Royal Society of London. Series B: Biological Sciences* **352**, 11 (1997).
- [14] A. Sasaki and Y. Haraguchi, Antigenic drift of viruses within a host: a finite site model with demographic stochasticity, *Journal of molecular evolution* **51**, 245 (2000).
- [15] M. Nowak and R. M. May, *Virus dynamics: mathematical principles of immunology and virology: mathematical principles of immunology and virology* (Oxford University Press, UK, 2000).
- [16] S. A. Kauffman *et al.*, *The origins of order: Self-organization and selection in evolution* (Oxford University Press, USA, 1993).
- [17] A. Korobeinikov, Global properties of basic virus dynamics models, *Bulletin of Mathematical Biology* **66**, 879 (2004).
- [18] R. R. Regoes, D. Wodarz, and M. A. Nowak, Virus dynamics: the effect of target cell limitation and immune responses on virus evolution, *Journal of theoretical biology* **191**, 451 (1998).
- [19] B. S. Adiwijaya, E. Herrmann, B. Hare, T. Kieffer, C. Lin, A. D. Kwong, V. Garg, J. C. Randle, C. Sarrazin, S. Zeuzem, *et al.*, A multi-variant, viral dynamic model of genotype 1 hcv to assess the in vivo evolution of protease-inhibitor resistant variants, *PLoS computational biology* **6**, e1000745 (2010).
- [20] V. Garcia, E. C. Glassberg, A. Harpak, and M. W. Feldman, Clonal interference can cause wavelet-like oscillations of multilocus linkage disequilibrium, *Journal of The Royal Society Interface* **15**, 20170921 (2018).
- [21] L. Yan, R. A. Neher, and B. I. Shraiman, Phylodynamic theory of persistence, extinction and speciation of rapidly adapting pathogens, *Elife* **8**, e44205 (2019).
- [22] T. Held, D. Klemmer, and M. Lässig, Survival of the simplest in microbial evolution, *Nature communications* **10**, 1 (2019).
- [23] D. A. Rasmussen and T. Stadler, Coupling adaptive molecular evolution to phylodynamics using fitness-dependent birth-death models, *Elife* **8**, e45562 (2019).
- [24] J. Marchi, M. Lässig, A. M. Walczak, and T. Mora, Antigenic waves of virus-immune coevolution, *Proceedings of the National Academy of Sciences* **118**, 10.1073/pnas.2103398118 (2021).
- [25] M. Lässig and M. Luksza, Adaptive evolution: Can we read the future from a tree?, *Elife* **3**, e05060 (2014).
- [26] R. A. Neher, C. A. Russell, and B. I. Shraiman, Predicting evolution from the shape of genealogical trees, *Elife* **3**, e03568 (2014).
- [27] M. M. Desai and D. S. Fisher, Beneficial mutation-selection balance and the effect of linkage on positive selection, *Genetics* **176**, 1759 (2007).
- [28] I. M. Rouzine, J. Wakeley, and J. M. Coffin, The solitary wave of asexual evolution, *Proceedings of the National Academy of Sciences* **100**, 587 (2003).
- [29] T. P. Velavan and C. G. Meyer, The covid-19 epidemic, *Tropical medicine & international health* **25**, 278 (2020).
- [30] Y. M. Bar-On, A. Flamholz, R. Phillips, and R. Milo, Science forum: Sars-cov-2 (covid-19) by the numbers, *elife* **9**, e57309 (2020).
- [31] P. V'kovski, A. Kratzel, S. Steiner, H. Stalder, and V. Thiel, Coronavirus biology and replication: implications for sars-cov-2, *Nature Reviews Microbiology* **19**, 155 (2021).
- [32] T. E. Harris *et al.*, *The theory of branching processes*, Vol. 6 (Springer Berlin, 1963).
- [33] R. Garcia-Millan, J. Pausch, B. Walter, and G. Pruessner, Field-theoretic approach to the universality of branching processes, *Physical Review E* **98**, 062107 (2018).
- [34] I. Pázsit and L. Pál, *Neutron fluctuations: A treatise on the physics of branching processes* (Elsevier, 2007).
- [35] M. Henkel, H. Hinrichsen, S. Lübeck, and M. Pleimling, *Non-equilibrium phase transitions*, Vol. 1 (Springer, 2008).
- [36] H. Hinrichsen, Non-equilibrium phase transitions, *Physica A: Statistical Mechanics and its Applications* **369**, 1 (2006).
- [37] T. E. Harris, Contact interactions on a lattice, *The Annals of Probability* **2**, 969 (1974).
- [38] H. Hinrichsen, Non-equilibrium critical phenomena and phase transitions into absorbing states, *Advances in physics* **49**, 815 (2000).
- [39] N. Friedman, S. Ito, B. A. Brinkman, M. Shimono, R. L. De Ville, K. A. Dahmen, J. M. Beggs, and T. C. Butler, Universal critical dynamics in high resolution neuronal avalanche data, *Physical review letters* **108**, 208102 (2012).
- [40] A. J. Fontenele, N. A. de Vasconcelos, T. Feliciano, L. A. Aguiar, C. Soares-Cunha, B. Coimbra, L. Dalla Porta, S. Ribeiro, A. J. Rodrigues, N. Sousa, *et al.*, Criticality between cortical states, *Physical review letters* **122**, 208101 (2019).
- [41] J. Pausch, R. Garcia-Millan, and G. Pruessner, Time-dependent branching processes: a model of oscillating neuronal avalanches, *Scientific Reports* **10**, 1 (2020).
- [42] P. Bak, C. Tang, and K. Wiesenfeld, Self-organized crit-

- icality: An explanation of the 1/f noise, *Physical review letters* **59**, 381 (1987).
- [43]S. S. Manna, Two-state model of self-organized criticality, *Journal of Physics A: Mathematical and General* **24**, L363 (1991).
- [44]S. Lübeck, Universal scaling behavior of non-equilibrium phase transitions, *International Journal of Modern Physics B* **18**, 3977 (2004).
- [45]C. Tang and P. Bak, Mean field theory of self-organized critical phenomena, *Journal of Statistical Physics* **51**, 797 (1988).
- [46]P. Alstrøm, Mean-field exponents for self-organized critical phenomena, *Physical Review A* **38**, 4905 (1988).
- [47]D. Dhar and S. Majumdar, Abelian sandpile model on the bethe lattice, *Journal of Physics A: Mathematical and General* **23**, 4333 (1990).
- [48]S. A. Janowsky and C. A. Laberge, Exact solutions for a mean-field abelian sandpile, *Journal of Physics A: Mathematical and General* **26**, L973 (1993).
- [49]R. García-Pelayo, I. Salazar, and W. C. Schieve, A branching process model for sand avalanches, *Journal of statistical physics* **72**, 167 (1993).
- [50]S. Zapperi, K. B. Lauritsen, and H. E. Stanley, Self-organized branching processes: mean-field theory for avalanches, *Physical review letters* **75**, 4071 (1995).
- [51]E. Ivashkevich, Critical behavior of the sandpile model as a self-organized branching process, *Physical review letters* **76**, 3368 (1996).
- [52]M. Katori and H. Kobayashi, Mean-field theory of avalanches in self-organized critical states, *Physica A: Statistical Mechanics and its Applications* **229**, 461 (1996).
- [53]S. di Santo, P. Villegas, R. Burioni, and M. A. Muñoz, Simple unified view of branching process statistics: Random walks in balanced logarithmic potentials, *Physical Review E* **95**, 032115 (2017).
- [54]K. B. Lauritsen, S. Zapperi, and H. E. Stanley, Self-organized branching processes: Avalanche models with dissipation, *Physical Review E* **54**, 2483 (1996).
- [55]D. B. Larremore, M. Y. Carpenter, E. Ott, and J. G. Restrepo, Statistical properties of avalanches in networks, *Physical Review E* **85**, 066131 (2012).
- [56]R. Otter, The multiplicative process, *The Annals of Mathematical Statistics* **20**, 206 (1949).
- [57]J. P. Sethna, K. A. Dahmen, and C. R. Myers, Crackling noise, *Nature* **410**, 242 (2001).
- [58]A. Baldassarri, F. Colaiori, and C. Castellano, Average shape of a fluctuation: Universality in excursions of stochastic processes, *Physical review letters* **90**, 060601 (2003).
- [59]O. Perković, K. Dahmen, and J. P. Sethna, Avalanches, barkhausen noise, and plain old criticality, *Physical review letters* **75**, 4528 (1995).
- [60]J. Wilting and V. Priesemann, Between perfectly critical and fully irregular: A reverberating model captures and predicts cortical spike propagation, *Cerebral Cortex* **29**, 2759 (2019).
- [61]P. Network, Pango network lineage database of sars-cov-2 (2021).
- [62]P. Network, Rules for the designation and naming of pango lineages (2021).
- [63]W. H. Organization, Who coronavirus (covid-19) database (2021).
- [64]A. Clauset, C. R. Shalizi, and M. E. Newman, Power-law distributions in empirical data, *SIAM review* **51**, 661 (2009).
- [65]Y. Virkar and A. Clauset, Power-law distributions in binned empirical data, *The Annals of Applied Statistics* **8**, 89 (2014).