Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
A theoretical exploration of the origin and early evolution of a pandemic

Yongsen Ruan, Haijun Wen, Xionglei He, Chung-I Wu *

State Key Laboratory of Biocontrol, School of Life Sciences, Sun Yat-sen University, Guangzhou 510275, China

Article history:
Received 18 July 2020
Received in revised form 15 September 2020
Accepted 3 November 2020
Available online 16 December 2020

Keywords:
SARS-CoV-2
COVID-19
Origin
Epidemics
Viral invasion
Herd immunity

1. Introduction

The pandemic coronavirus disease 2019 (COVID-19) is caused by the virus severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2). By now, the number of scientific reports has been astounding, covering almost all aspects of COVID-19 [1–16]. Among the questions, the origin may be the most intriguing and controversial one. Where, when, and how did the pandemic originate? Although these questions are flooding the public domains, few seem to realize how loaded these questions are. The popular concept of the origin of SARS-CoV-2 is a non-evolutionary one. The idea is to search for animals that harbor the human SARS-CoV-2; a successful find would then define the place of origin. Here, we propose such a model whereby evolution occurs in both the virus and the hosts (where the evolution is somatic; i.e., in the immune system). The hosts comprise three groups – the wild animal hosts, the nearby human population, and farther-away human populations. The theory suggests that the conditions under which the pandemic has initially evolved are: (i) an abundance of wild animals in the place of origin (PL0); (ii) a nearby human population of low density; (iii) frequent and long-term animal-human contacts to permit step-by-step evolution; and (iv) a level of herd immunity in the animal and human hosts. In this model, the evolving virus may have regularly spread out of PL0 although such invasions often fail, leaving sporadic cases of early infections. The place of the first epidemic (PL1), where humans are immunologically naïve to the virus, is likely a distance away from PL0. Finally, this current model is only a first attempt and more theoretical models can be expected to guide the search for the origin of SARS-CoV-2.

* Corresponding author.
E-mail address: wzhongyi@mail.sysu.edu.cn (C-I Wu).

https://doi.org/10.1016/j.scib.2020.12.020
2095-9273/© 2020 Science China Press. Published by Elsevier B.V. and Science China Press. All rights reserved.
synonymous changes in the receptor-binding domain (RBD) of the S protein vis-à-vis their putative ancestor found in the wild animals [27,30].

In this view, the “origin of SARS-CoV-2” should be an evolutionary process that would take some time to complete. In fact, there may be several distinct evolutionary processes—the evolution of the viral genome in wild animal hosts and the further evolution in humans. Furthermore, the immune responses in the animal hosts and humans add further complexities. The development of immunity can be considered somatic evolution of the immune system [31,32]. These processes happen on a short time scale relative to the evolutionary processes of the germline.

Finally, a companion study by Ruan et al. [33] on the founder population size concludes that outbreaks are often started by 5–10 infected travelers. Hence, once SARS-CoV-2 has evolved in its place of origin (PL0 for short), the fully evolved strain may easily spread elsewhere by a few travelers. The epidemics erupt outside of (rather than within) PL0 because the outside populations have not gained the host immunity.

2. The outline of the model

Fig. 1 is a model tracing the early evolution of SARS-CoV-2 in PL0 that eventually unleashes the global pandemic. As stated, the model, although a strictly theoretical construct, is based on observations reported for earlier epidemics of HIV, influenza A, and SARS [23–28]. The terms and symbols central to this model are summarized in Table 1. In PL0, there should be an abundance of wild animals in which the virus has evolved step-by-step over a prolonged period of time. Such a place, possibly a wildlife reserve or a remote countryside, might generally have a local human population (referred to as H0) of low density.

We model the infectiousness of the virus in both the animal hosts and humans. As in the companion study [33], the infectiousness is expressed by the distribution of k, which is the number of individuals a carrier will infect in a time period (see below). The population dynamics of the virus depends on the distribution of k, in particular, its mean (E(k)) and variance (V(k)). E(k) is closely related, but not identical, to the key epidemiological parameter R0 [33]. The population size of the virus at time t is N(t), defined as the number of infected individuals in the host population.

![Fig. 1. A model for the origin and early evolution of the epidemics. (a, b) The evolution of the virus in the place of origin (PL0; inside the green box). (c) The invasion outside of PL0.](image-url)
Fig. 1a depicts the evolution of the virus in animal hosts (referred to as $A_0$). Here, the hosts may be a species (say, of bats) or more than one species of wild animals among which the virus circulates. The infectiousness is expressed as $E_0(k)$ where $A$ denotes the wild animals. $E_0(k)$ would increase when the virus acquires an adaptive mutation and would decrease as the host develops the herd immunity. The process depicted in Fig. 1a is in the evolutionary time scale and only the most recent events are shown. Also, $E_0(k)$ is >1 as a virus with $E_0(k) \leq 1$ would be quickly eliminated (see below).

In Fig. 1b, the spread of the virus from the animal hosts to $H_0$ is depicted. As the virus evolves, it may have invaded $H_0$ multiple times (indicated by thin arrows) but failed to trigger an epidemic. Because natural selection has been continuously working in the animal hosts (but not in humans), higher infectiousness in the wild animal hosts than in $H_0$ is expected. Hence, $E_0(k) > 1$ is rare and failed invasions should be common ($E_0(k)$ denotes the viral infectiousness in $H_0$). The failed invasions may nevertheless cause the gradual build-up of herd immunity in $H_0$. Thus, the trend of $E_0(k)$ in $H_0$ goes down, similar to that in the animal hosts.

The crucial step for the virus to infect humans is the emergence of a proto-human virus, as shown by the thick arrow in the middle of Fig. 1b. This proto-human strain, $V_0$, has an $E_0(k) > 1$ albeit unlikely to be much larger. For that reason, $V_0$ may sustain the infection in $H_0$ long enough to acquire new mutations for further enhancement of infectiousness in humans. The duration of the infection and the total number of infections will be the crux of the theory developed of infectiousness in humans. The duration of the infection and the gradual build-up of herd immunity in $H_0$ goes down, similar to that in the animal hosts.

Finally, given the small number of infections needed to start an outbreak of COVID-19 [33], it would be inevitable that a global pandemic will ensue (big solid arrow, Fig. 1c).

3. Results

3.1. Model predictions of the long-term co-existence of virus and hosts

We track the viral population size, $N(t)$, as the virus evolves to cope with the herd immunity of the hosts. $N(0)$ is the viral population size at the time of invasion and is equivalent to $I_0$ of Ruan et al.'s [33] study. Since every infected individual is assumed to have identical behavior, we present the results for $N(0) = 1$. For a viral invasion, the probability of ultimate extinction is denoted by $\lim P[N(t) = 0]$. If the extinction probability is $u$ with $N(0) = 1$ (i.e., $u = \lim P[N(t) = 0|N(0) = 1]$), then the probability would be $u^t$ when $N(0) = n$. In humans, the viral generation time is assumed to be 4 days [5] as in Ruan et al.'s study [33].

3.1.1. A partial model with constant $E(k)$ when there is no host immunity

We first consider a partial model with neither host immunity nor viral evolution and, thus, with constant $E(k)$. Viral invasion of the host population would fail when all infected individuals fail to infect others (i.e., $k = 0$) in any generation. This could happen if $E(k)$ is smaller, or at least close to 1. In Table 2 (also Tables S1–S4 online), we show the probability of ultimate extinction ($u$) for $N(0) = 1$ under various distributions of $k$ (power law, Poisson, and binomial). Analytically, $u$ is the smallest non-negative root of the following equation:

$$G_u(u) = \sum_{k=0}^{\infty} p_k u^k = u,$$

(1)

where $p_k$ is the probability distribution of $k$. If $k$ follows the Poisson distribution, we can obtain its extinction probability by the following equation:

$$G_u(u) = e^{-(a-1)} = u.$$

Hence,

$$u = \frac{\text{LambertW}(-ae^{-a})}{a},$$

(2)

where Lambert $W$ is the Lambert $W$ function [34]. The detailed derivations for the power law distribution are shown in the Supplementary information online.

Eq. (1) shows the dependence of the extinction probability on the distribution of $k$. In this study, $k$ follows three different distributions (power law, Poisson, and binomial distribution) with $E(k)$ ranging from 1 to 5. As shown in Table 2, the probability of extinction is very high if $N(0) = 1$ under the power law distribution. If $E(k) = 2$, 73% of the invasions would fail. Even if $E(k) = 4.5$ as is observed in COVID-19, the chance of failure is still around 43%. In infections with $E(k) < 2$, the probability of failed infections would be higher than 90%. Note that, given the same $E(k)$, the chance of failure is generally lower if $V(k)$ decreases. For example, if $V(k) = E(k)$ as in the Poisson distribution, the probability of failure with $E(k) = 4.5$ would only be about 1%.

Fig. 2a–c show 50 replicates of the changes in $N(t)$ as a function of time when $E(k)$ ranges between 1.05 and 1.5. It appears that the population would either go extinct in the first few generations or escape extinction by continuing to grow in size. In particular, when $N(t)$ reaches 100, a successful epidemic would be certain. Thus, the virus would either disappear or cause an epidemic in this non-evolutionary model.

3.1.2. The full model with host immunity and continual viral evolution

As the virus and the host cannot coexist in the long run under constant $E(k)$, we consider a more realistic scenario in which the host would develop immunity to suppress the virus and the viral genome would evolve to evade the suppression. This scenario would be analogous to an evolutionary arms race, but note that the (somatic) evolution in the host happens in the immune system [31,32]. For that reason, the arms race has the dynamics much faster than the germline coevolution [35–37].

In modeling this arms race, the host would repress the $E(k)$ of the virus to below 1 and the virus may acquire adaptive mutations to drive $E(k)$ above 1. In the previous study [33], we let the distribution of $k$ follow different distributions, in particular, the Poisson distribution with $V(k) = E(k)$ and the power law distribution (in this case, we let $V(k) = SE(k)$). Unless held back by a new adaptive mutation in the viral genome, we assume that the herd immunity will decrease $E(k)$ as specified below:

$$E(k; t) = E(0) - \frac{AE}{T} t^2,$$

(3)

where $E(0)$ is the initial $E(k)$, $E(k; t)$ is the $E(k)$ at generation $t$, $T$ is the time required to reduce $E(k)$ by $\Delta E$. The parameter $z$ determines the curvature of $E(k)$ over time as shown in Fig. 3a. When $z = 1$, $E(k)$
declines linearly over time. When \( z > 1 \), \( E(k) \) declines slowly in the beginning but the decline gradually speeds up yielding a convex shape as shown in red in Fig. 3a. This pattern may be seen when the virus invades an immunologically naïve population. In such a population, the herd immunity slowly builds up but accelerates later on. The opposite is seen with \( z < 1 \) whereby the herd immunity is high in the beginning and \( E(k) \) decreases rapidly. The decline in \( E(k) \) slows down when most individuals in the population have acquired immunity, hence giving rise to the concave appearance shown in blue in Fig. 3a. This pattern may apply to the repeated failed invasions of the virus as the later rounds of invasions would encounter strong immunity as they enter the host population.

For each new adaptive mutation, the \( E(k) \) will increase as follows:

\[
E(k)_{+1} = E(k) + J,
\]

(4)
where \( J \) is infectivity gain. Unless otherwise specified, we assume \( J \) follows the exponential distribution with mean equal to \( m \). Here, we assume the accumulation of adaptive mutations follows the Poisson process with the average rate of \( U \). For a long-term persistence in animal (or human) hosts, the decline of \( E(k) \) by herd immunity should be held back by a new adaptive mutation. During the time gap waiting for a new adaptive mutation (\(-1/U\)), the decline of \( E(k) \) is \( m \) approximately. A simple way to satisfy this condition is to let \( T = 1/U \), \( \Delta E = m \). In this case, the expected infectivity of the virus in the hosts (\( A_0 \) or \( H_0 \)) is now a function of time.

\[
E(k; t) = E(0) - q\Delta E - \frac{\Delta E}{T}(t - qT) + \sum_{k=1}^{M_t} J_k, \quad T = 1/U, 
\]

where \( M_t \) is the number of adaptive mutation at time \( t \), \( J_k \) is the infectivity gain of a particular adaptive mutation, and \( q \) is the quotient (i.e., the integer part of the ratio of \( t \) to \( T \)). Note that the \( E(k; t) \) value in humans (i.e., \( E_h(k) \)) is much lower than the corresponding \( E(k; t) \) in wild animals (i.e., \( E_w(k) \)) because the infectiousness \( E(k) \) and \( V(k) \) evolves in the wild animals. For example, the viral RBD possesses an exceptionally high affinity for angiotensin-converting enzyme 2 (ACE2) only in the species where the virus is found originally [27,30,38,39].

When the host develops the herd immunity against the viral strain that has a non-zero probability of long-term persistence (i.e., \( E(k) > 1 \)), those persistent cases would now go extinct eventually. However, the infection would last much longer than the invasions with \( E(k) < 1 \). With an initial population size \( N(0) = 1 \), we denote the duration between the initial infection and extinction as \( T_{inf} \). The total number of infected individuals summed over all generations is designated as \( N_{inf} \):

\[
N_{inf} = \sum_{t=0}^{T_{inf}} N(t), 
\]

where \( N(t) \) is the infected cases at generation \( t \). Note \( N(t) \) is never greater than the population size of \( H_0 \), but \( N_{inf} \) can be larger since it is a cumulative count. \( N_{inf} \), positively correlated with \( T_{inf} \), would determine the probability of new adaptive mutations in the viral strain, which will then overcome the herd immunity.

### 3.2. Modeling the evolution in the wild animals of \( P_{L0} \)

For the evolution of the virus in the animal host as sketched in Fig. 1a, their long-term existence is likely the outcome of an evolutionary arms race. Fig. 3b shows an example where each viral mutation that increases the infectiousness is countered by the immune response of the host. Each response by the host is further countered by another viral mutation that increases the infectiousness (Eq. (5)). The viral population thus waxes and wanes over a long time span (Fig. 3c). The example is only a demonstration that realistic parameter values can indeed lead to the long-term coevolution between the virus and the immune system of the host. The parameter space conducive for co-existence is large, especially when \( V(k) \) decreases much faster than \( E(k) \). We should add that a simple condition for the long-term existence of virus is for \( E(k) \) to approach 1 and \( V(k) \) to approach 0 as \( t \) increases. In other words, the number of infected individuals in the population would stay constant over generations, thus giving the virus plenty of time to acquire new adaptive mutations.

### 3.3. Modeling the viral evolution in \( H_0 \) at \( P_{L0} \)

The main challenge is to explain how a highly infectious strain like SARS-CoV-2 could have evolved in humans. Given that humans are expected to develop immunity as the virus evolves step by step, neither side should have such an overwhelming advantage over the other in the arms race. In contrast, human populations elsewhere are immunologically naive; hence, the high infectiousness is plausible outside of \( P_{L0} \).

In \( P_{L0} \), the virus would likely infect the local human population regularly. However, the infection is usually unsustainable since the contagion has been selected in the animal hosts, not in humans. The failed invasions nevertheless could elicit a degree of herd immunity in \( H_0 \) [40]. A crucial event would then be the emergence of a rare strain that happens to be moderately infectious in humans, referred to as \( V_0 \). It would seem unlikely that \( V_0 \) could be so infectious as to trigger a global pandemic in humans. Instead, we assume that \( V_0 \) only needs to sustain the infection long enough to acquire advantageous mutations for further evolution in human populations.

In Fig. 4, we show the invasion of \( V_0 \) with a very modest \( E(k) \) of 1.1. Such a strain would succeed in the invasion with a probability of <5% (Table 2). Furthermore, when the host population evolves the immunity, \( E(k) \) would drop below 1, leading to the virus’s eventual extinction. Fig. 4 presents the distribution of \( T_{inf} \) and \( N_{inf} \) in the process of invasion-to-extinction.

Fig. 4a is most interesting as \( T_{inf} \) and \( N_{inf} \) both show a bi-modal distribution. \( V_0 \) either dies out by generation 30 or sustains the invasion for >130 generations. The reason for this bimodal distribution is that ~95% of the invasions with \( E(k) = 1.1 \) would fail, which corresponds to the rapidly failed cases. The remaining 5% would become extinct when the host population develops herd immunity, which takes another 100 generations in this simulation. In this process, \( N_{inf} \) would be either <100 or >60,000. Only in the latter cases would \( N_{inf} \) be large enough to yield adaptive mutations for the virus to overcome the immune suppression (see Supplementary information online for the probability of acquiring mutations). The fourth panel of Fig. 4a shows that, for about 50 generations, \( V_0 \) is found in almost every individual in \( H_0 \).

While Fig. 4a shows the infection of \( V_0 \) has only a 5% chance of sustaining the infection, the chance by the 10th infection would be 40% (~1-0.9510). However, since the immunity would develop gradually, this simple calculation may not be valid. In later invasions, the probability of failure could become higher. With this consideration, the virus would generally disappear by generation 50 (Fig. 4b) and, in later infections, the invasion may all fail by generation 10 (Fig. 4c). In other words, \( V_0 \) may have few chances to invade the human population. If the invasion fails, the human host would be immunologically vigilant, thus thwarting subsequent attempts. Although we let \( k \) follow power law distribution with \( V(k) = 5^k(k/100-1) \) here, the same pattern can be obtained with larger or smaller \( V(k) \) (Figs. S1 and S2 online).

In this model, the arms race between the invading virus and the human host is a long-drawn-out battle in \( P_{L0} \). Each new mutation that may allow the virus to invade the human population from the animal hosts would usually fail. The failed virus would need another new mutation to start another round of battle. The biological characteristics of \( P_{L0} \)s are, therefore, (i) a place with a high density of wild animals (in particular, bats), (ii) a local low-density human population (such that the invasions usually fail), (iii) long-term coexistence and frequent contacts between the animal and human populations, and (iv) a level of herd immunity in the animal and human hosts. All these conditions are conducive for the virus to accumulate many mutations step-by-step toward the high infectiousness necessary for the pandemic outside of \( P_{L0} \).

### 3.4. The spread of \( V_1 \) from \( P_{L0} \) to \( P_{L1} \) — the first epidemic

The conditions in \( P_{L0} \) should allow a \( V_0 \) strain to sustain itself in \( H_0 \) long enough to acquire beneficial mutations that are adaptive specifically in humans. For simplicity, we assume that a \( V_1 \) strain
carrying the first beneficial mutation is capable of triggering an epidemic (if \( V_1 \) fails to establish itself, the next adaptive mutation would create a \( V_2 \) strain and, if it fails too, there may be a \( V_3 \) and so on). Note that \( V_1 \) would be repressed by the cumulative herd immunity in \( H_0 \) but \( V_1 \) should have a chance of spreading outside of \( P_{L0} \).

In Fig. 1c, once \( V_1 \) succeeds in spreading to an immunologically naïve population in \( P_{L1} \), the pandemic would be a near certainty since no population outside of \( P_{L0} \) is immune to the virus. Ruan et al.’s study \[33\] shows that the epidemics in many countries are initiated by 5–10 travelers into the region. Hence, the spread to \( P_{L1} \) should be likely once SARS-2-CoV-2 is fully evolved in \( P_{L0} \). With \( E(k) = 4.5 \) and 5 carriers, Table 2 shows the probability of invasion failure to be < 2% (\( 0.45^5 = 1.8\% \)).

4. Discussion

In light of the previous models for viral epidemics \[18–20,22\], we propose the current model which has several key features: First, viral adaptation to the new human host requires many evolutionary steps. This assumption has been validated for other viral pathogens \[23–28,30\]. Previous models have also pointed out that cross-species transmissions evolve mainly over the time scale of millions of years \[21\]. Thus, host-switching like other complex adaptations is a multi-step process. Second, the onset of an epidemic is highly unpredictable \[18\] and this current model has built a mathematical framework to address the stochasticity. Third, the multi-step evolution often leads to a stasis between the host and the pathogen, rather than a global pandemic. The current model proposes to separate the place of origin and places of epidemics by geography.

As the virus evolves, failed invasions prior to the first epidemic in \( P_{L1} \) seem likely (blue arrows of Figs. 1b and 2). We shall refer to places of possible earlier invasions as \( P_{Lx} \) where sporadic cases of COVID-19 may be found. Such failed invasions would be most informative about the early stages of the pandemic, if the genomic sequences can be obtained and examined. There have been multiple reports of such a possibility in France \[41\], Japan, and various parts of the US. These reports have often been discredited without further investigations that can definitively rule in, or rule out, SARS-CoV-2’s involvement. In particular, patterns of traveling to \( P_{Lx} \)’s may offer a clue of \( P_{L0} \), where the local human population should show a degree of immunity to SARS-CoV-2. Some researchers \[40\] appear to suggest candidate \( P_{L0} \) sites to be where humans come in frequent contact with bats. Given the existence of a large reservoir of coronaviruses in wild bats, people in the countryside with a sizable bat population may show some immunity to the viruses \[1,40\]. At present, the closest strain isolated in wild animals is 10% divergent from the genome of SARS-CoV-2 \[6,42\], a distance estimated to take 30–300 years to evolve (Jian Lu, unpublished results). At present, we are still far from finding the origin of this pandemic.

In the popular view, the “first” SARS-CoV-2 corresponds to one of the strains circulating in humans. An extension of this view is that the place of origin is the same as \( P_{L1} \). This extension is problematic. In the Guns, Germs and Steel, Diamond \[43\] has documented many instances that pathogens spread to the places where they encounter low resistance (such as smallpox in the Aztec populations). The geographical separation between the place of origin of any taxon (not just virus) and the site of proliferation has led Gould et al. to the explanation of “punctuated equilibrium” \[44\]. The fossil records in the place of origin may be continual but few; in contrast, their abundance at the site of the outbreak would appear to suddenly come out of nowhere. For a recent example, the widely discussed “Spanish flu” of 1918 broke out in the war zone but most likely originated in Kansas \[45,46\]. In short, if \( P_{L0} \) is the same as \( P_{L1} \), which is widely assumed, then the origin and evol-
tion of SARS-CoV-2 must be very unusual. A theory for such an unusual origin is thus needed.

Like many other evolutionary questions on the origin, where, when, and how SARS-CoV-2 becomes fully evolved will remain an intriguing question. We suggest the question be about the early evolution of SARS-CoV-2, rather than about the “origin”. The former implies a process whereas the latter seems to mean a single time point. This distinction is important as seen in the earlier debates on the “origin” of dogs [47] and new species in environments [48].

Conflict of interest

The authors declare that they have no conflict of interest.

Acknowledgments

This work was supported by the National Natural Science Foundation of China (31730046, 91731000, 31900417, and 81972691), Guangdong Basic and Applied Basic Research Foundation (2020B1515020030, 2019A1515010708), and the National Key Research and Development Project of China (2020YFC0847000). We thank Drs. Yaping Zhang, Jian Lu, Xuemei Lu, Suhua Shi, Weiiwei Zhai, Jianzhi George Zhang, Bingjie Chen, Zheng Hu, and Jindong Zhao for comments and suggestions.

Author contributions

Chung-I Wu and Yongsen Ruan designed the study and constructed the theoretical framework. Yongsen Ruan did the simulation with the help of Chung-I Wu. Yongsen Ruan and Chung-I Wu wrote the manuscript. All authors interpreted the findings, revised the manuscript, and approved the final version for publication.

Appendix A. Supplementary materials

Supplementary materials to this article can be found online at https://doi.org/10.1016/j.scib.2020.12.020.

References

[1] Andersen KG, Rambaut A, Lipkin WI, et al. The proximal origin of SARS-CoV-2. Nat Med 2020;26:457–9.
[2] Ferrer I, Wymant C, Kendall M, et al. Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing. Science 2020;368: eabb6936.
[3] Li R, Pei S, Chen B, et al. Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS-CoV-2). Science 2020;368:489–93.
[4] Liu Y, Gayle AA, Wilder-Smith A, et al. The reproductive number of COVID-19 is higher compared to SARS coronavirus. J Travel Med 2020;27:e1aa021.
[5] Wöllel R, Corman VM, Gugmemos et, al. Virological assessment of hospitalized patients with COVID-2019. Nature 2020;581:465–9.
[6] Zhang T, Wu Q, Zhang Z. Probable pangolin origin of SARS-CoV-2 associated with the COVID-19 outbreak. Curr Biol 2020;30:1346–1351 e2.
[7] Zhou P, Yang XL, Wang XG, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature 2020;579:270–3.
[8] Boni MF, Lemey P, Jiang X, et al. Evolutionary origins of the SARS-CoV-2 sarbecovirus lineage responsible for the COVID-19 pandemic. Nat Microbiol 2020;5:1408–17.
[9] Liu YX, Zhang C, Huang FM, et al. Elevated plasma levels of selective cytokines in COVID-19 patients reflect viral load and lung injury. Nat Sci Rev 2020;7:1003–11.
[10] Yang XL, Wu CC, Li X, et al. On the origin and continuing evolution of SARS-CoV-2. Nat Sci Rev 2020;7:1012–23.
[11] Koerber B, Fischer WM, Gnanaakaran S. Tracking changes in SARS-CoV-2 spike: evidence that D614G increases infectivity of the COVID-19 virus. Cell Host Microbe 2020;28:192–202.
[12] Wu CI, Poo MM. Moral imperative for immediate release of 2019-nCoV sequence data. Nat Sci Rev 2020;7:719–20.
Yongsen Ruan, Postdoctoral Fellow, School of Life Sciences, Sun Yat-sen University. He is interested in theoretical population genetics and specialized in computational biology and mathematical modeling.

Chung-I Wu, Professor, School of Life Sciences, Sun Yat-sen University. He is an evolutionary biologist and a member of the Academia Sinica. For the past 30 years, he has been a professor at the University of Chicago and the head of the Department of Ecology and Evolution (1998–2008). From 2008 to 2014, he served as the director of the Beijing Institute of Genomics of the Chinese Academy of Sciences. In 2004, he cooperated with Guoping Zhao and other domestic experts to reveal the evolutionary dynamics of the SARS virus. He has also accomplished a series of work on the origin and evolution of SARS-CoV-2.