

## Unravelling the origin of SARS-CoV-2: is the model good?

R. Frutos<sup>1</sup>, L. Gavotte<sup>2</sup> and C. A. Devaux<sup>3</sup>

1) Cirad, UMR 17, Intertryp, 2) Espace-Dev, Université de Montpellier, Montpellier and 3) IHU-Méditerranée Infection and CNRS, Marseille, France

### Abstract

MacLean and colleagues recently published a very elegant analysis demonstrating that SARS-CoV-2 carries signs of positive selection and that it was already adapted to humans prior to the emergence of COVID-19. Using the Spillover theory as a reference model for zoonotic emergence, they conclude that SARS-CoV-2 must have acquired this human adaptation in bats. We reinterpreted the data from MacLean et al. using a different model of zoonotic emergence as reference, the Circulation model. The use of the Circulation model provides a more parsimonious interpretation showing that this adaptation to humans occurs in the human population after primo infection.

© 2021 The Author(s). Published by Elsevier Ltd.

**Keywords:** Covid-19, circulation model, emergence, evolution in humans, SARS-CoV-2

**Original Submission:** 9 May 2021; **Revised Submission:** 12 June 2021; **Accepted:** 9 July 2021

**Article published online:** 16 July 2021

**Corresponding author:** Roger Frutos, Intertryp – UMR 17, CIRAD, TA A17/G Campus International de Baillarguet, 34398, Montpellier Cedex 5, France  
**E-mail:** [roger.frutos@cirad.fr](mailto:roger.frutos@cirad.fr)

Comments on the article “Natural selection in the evolution of SARS-CoV-2 in bats created a generalist virus and highly capable human pathogen” published in PLoS Biology by MacLean et al. DOI: <https://doi.org/10.1371/journal.pbio.3001115> [1].

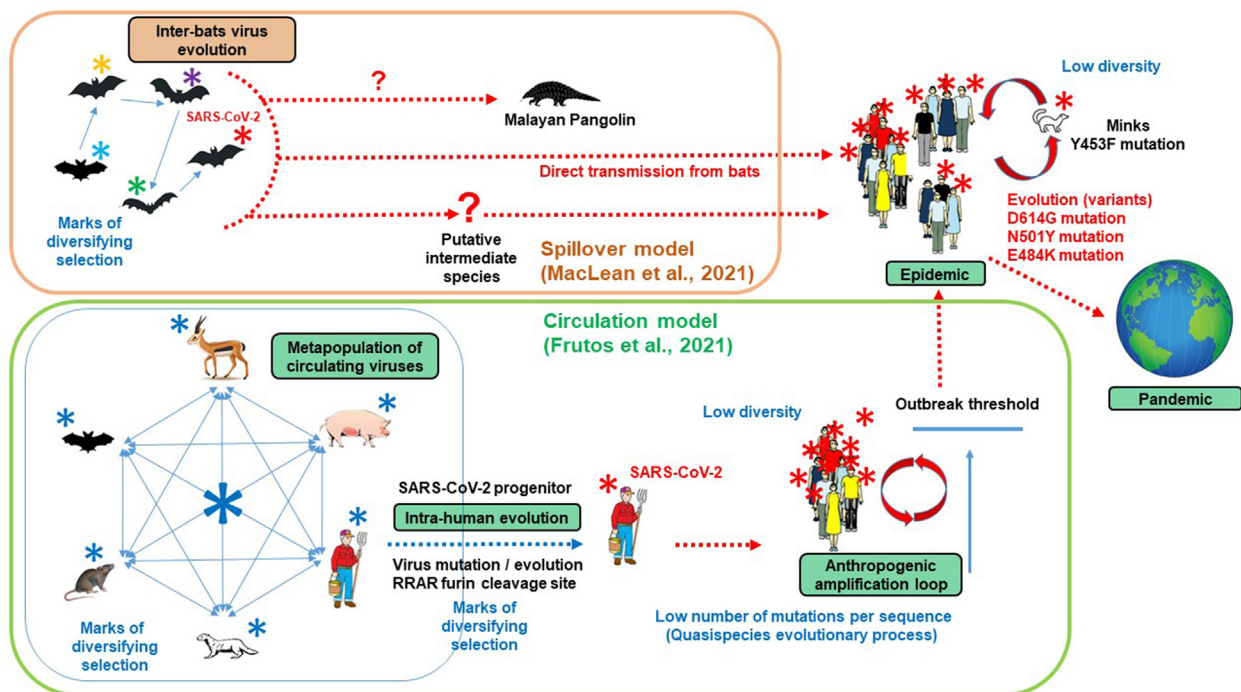
In this article, Maclean and colleagues highlighted that “SARS-CoV-2 has apparently required little to no significant adaptation to humans since the start of the Coronavirus Disease 2019 (COVID-19) pandemic and to October 2020”. Based on various bioinformatic

analyses, they observed that “purifying selection is much weaker in SARS-CoV-2 than in related bat Sarbecoviruses”. They also found “evidence for significant positive episodic diversifying selection acting at the base of the bat virus lineage SARS-CoV-2 emerged from, accompanied by an adaptive depletion in CpG composition presumed to be linked to the action of antiviral mechanisms in these ancestral bat hosts”. A major conclusion from MacLean and colleagues is that traits essential for effective human infection have been selected in bats long before the emergence of COVID-19. They also concluded that the circulation of such a virus with “generalist property” in the bat reservoir led to a virus highly capable of “readily transmit to other animals”.

Our comment is by no means a criticism of this article, which we find excellent with solid and appropriate methods, and thorough and relevant analyses. However, we believe that the authors used an explanatory model for the emergence of zoonotic diseases (i.e. the “spillover model”) that is now found to be unable to describe the emergence of COVID-19. This model has led to a misinterpretation of otherwise solid and convincing results. The spillover model specifies that the human-adapted virus must be present prior to human infection, and human competency was thus already present before the COVID-19 outbreak. Under this model, and owing to the fact that no intermediate species was ever found until now, MacLean and colleagues logically concluded that SARS-CoV-2 must have been selected in bats as an already human-competent virus. The authors themselves seem to have felt these problems and several times had to resort to assumptions.

We recently proposed an alternative to the spillover model for explaining the emergence of SARS-CoV-2: the circulation model [2,3]. The circulation model states that the virus causing the disease in human populations does not exist yet in the wild. It evolves in humans following primo-infection by a circulating proximal virus. Viruses belonging to the same metapopulation circulate at a low level in the wild simply upon contact and minimal compatibility. The same happens at the animal-human interface. Following infection of a new host, e.g. humans, the virus evolves and adapt better to this new host. However, this is not sufficient to trigger an epidemic in humans. Following this phase of circulation at a low level in the human population, societal events must occur, e.g. meetings, gatherings, markets, which allow the amplification of the viral population to the point where the epidemic threshold is reached and the epidemic starts.

We suggest here to carry a new reading and interpretation of the results from MacLean et al. [1] according to the circulation model. MacLean and colleagues compared SARS-CoV-2 sequences, thus post-epidemic sequences of the human virus, and coronavirus sequences from *Rhinolophus affinis* and *Rhinolophus malayanus* bats, i.e. RaTG13 and RmYN02 [4,5]. Other



**FIG. 1.** Representation of the conclusions from MacLean and colleagues according to the spillover and circulation models. The figure is adapted from MacLean et al., 2021 [1] and Frutos et al., 2021 [3]. The text in blue corresponds to the observations from MacLean et al., 2021 [1]. Arrows, viruses and text in red correspond to events involving SARS-CoV-2. Under the spillover model, the virus found in pangolins must come from another source than SARS-CoV-2. This is represented by the question mark. Under the circulation model, the species shown in the figure symbolize the circulation of the virus metapopulation in various animal species. They do not specifically represent species in which SARS-CoV-2 has been found.

SARS-CoV-2-related viruses from *Rhinolophus* bats have since then been described from Cambodia and Thailand [6,7], indicating that this group of bat viruses was also present outside China. MacLean and colleagues found evidence of several positive selection sites in addition to human-competency traits. Using the recently described circulation model as a reference to revisit MacLean and colleagues' data, we come to quite a different interpretation. The circulation model states that there is no reservoir and no intermediate species but a circulation of viruses from one susceptible host to another through contact. This corresponds to the conclusion by MacLean and colleagues that the SARS-CoV-2 proximal virus must have circulated in different hosts owing to the presence of positively selected sites. Indeed, they state that "Our analysis finds that diversifying selection left its imprints primarily in the deepest branches of the nCoV clade or lineage leading to it, with no evidence of selection in the terminal branch leading to SARS-CoV-2. This is consistent with the nonhuman progenitor of SARS-CoV-2 requiring little or no novel adaptation to successfully infect humans". However, a major difference is that under the circulation model there is no obligation to maintain this circulation exclusively among bat species because no intermediate species could be identified. Under the circulation model, a virus metapopulation is circulating in

different susceptible host species, including humans, and is evolving in-host through a quasispecies process. SARS-CoV-2 is the result of this evolutionary process in humans. This leads to a different evolution in each host species, and thus, to a range of related but still different viruses. This is what we observe. This quasispecies process of evolution also corresponds to some observations from MacLean and colleagues, i.e. "The vast majority of 20,687 observed mutations occur at very low frequency, with 79% of mutations observed in 10 or fewer of the 133,741 SARS-CoV-2 genome sequences analysed". This is what is expected to occur in a quasispecies evolutionary process. In the circulation model, the virus is already present in humans before acquiring through in-host evolution the specific traits of the virus causing the pandemic. This is why it cannot be found in any intermediate species and why there is indeed no intermediate species. A virus proximal to SARS-CoV-2 was already circulating in humans. In their study, MacLean and colleagues compared post epidemic SARS-CoV-2 sequences to bat betacoronaviruses sequences, but the genetic distance between the two sets of samples is too big, and more importantly, could not explain the evolutionary features detected by the authors making the evolution of SARS-CoV-2 in bats to be the only possible conclusion. However, SARS-CoV-2 was never

found in any wild animal, only in domestic (cats), captive animals (minks, tigers, etc.) infected by humans. SARS-CoV-2 was also found in wild animals in contact with captive specimens of the same species (minks). The authors report several times in their article that the SARS-CoV-2 genome is stable in humans. This is normal since all human viruses analysed were exclusively obtained post-pandemic (i.e. after the selection process leading to an epidemic virus according to the circulation model). They say, for instance: *“Even in Spike, which is being assiduously scrutinized for selection due to its immunogenic and phenotypic importance, overall selective pressure is stable over time and consistent with weak purifying selection. The genetic homogeneity of SARS-CoV-2 results in very shallow phylogenetic trees, despite > 130,000 collected sequences, with cumulative branch lengths only about 0.6 (S) and 0.4 (RdRp) substitutions/site”*. This is expected at this stage since the intra-human evolution already took place prior to reaching the outbreak threshold and to the beginning of the epidemic/pandemic [3].

We propose in Fig. 1 a positioning of the conclusions and observations MacLean and colleagues on both the spillover and circulation models.

As a conclusion, the analyses conducted by MacLean and colleagues are very good and meaningful and provide good support to the emergence of SARS-CoV-2 according to the circulation model with which they fit well.

## Conflict of interest

The authors declare no conflict of interest.

## References

- [1] MacLean OA, Lytras S, Weaver S, Singer JB, Boni MF, et al. Natural selection in the evolution of SARS-CoV-2 in bats created a generalist virus and highly capable human pathogen. *PLoS Biol* 2021;19:e3001115.
- [2] Frutos R, Serra-Cobo J, Chen T, Devaux CA. COVID-19: time to exonerate the pangolin from the transmission of SARS-CoV-2 to humans. *Infection. Genet Evol* 2020;84:104493.
- [3] Frutos R, Gavotte L, Devaux CA. Understanding the origin of COVID-19 requires to change the paradigm on zoonotic emergence from the spillover model to the viral circulation model. *Infection. Genet Evol* 2021;104812. <https://doi.org/10.1016/j.meegid.2021.104812>.
- [4] Zhou P, Yang XL, Wang XG, Hu B, Zhang L, et al. A pneumonia outbreak associated with a new 654 coronavirus of probable bat origin. *Nature* 2020;579:270–3.
- [5] Zhou H, Chen X, Hu T, Li J, Song H, et al. A novel bat coronavirus closely related to SARS-CoV-2 contains natural insertions at the S1/S2 cleavage site of the spike protein. *Curr Biol* 2020;30:2196–203. e3.
- [6] Hul V, Delaune D, Karlsson EA, Hassanin A, Tey PO, et al. A novel SARS-CoV-2 related coronavirus in bats from Cambodia. *BioRxiv* 2021. <https://doi.org/10.1101/2021.01.26.428212>.
- [7] Wacharapluesadee S, Tan CW, Maneeorn P, Duengkhae P, Zhu F, et al. Evidence for SARS-CoV-2 related coronaviruses circulating in bats and pangolins in Southeast Asia. *Nature Commun* 2021;12:1–9.