
Abstract

Evolutionary origins of the SARS-CoV-2 sarbecovirus lineage responsible for the COVID-19 pandemic

Maciej F. Boni, Philippe Lemey, Xiaowei Jiang, Tommy Tsan-Yuk Lam, Blair W. Perry, Todd A. Castoe, Andrew Rambaut & David L. Robertson

***Nature Microbiology* (2020)**

There are outstanding evolutionary questions on the recent emergence of human coronavirus SARS-CoV-2 including the role of reservoir species, the role of recombination and its time of divergence from animal viruses. We find that the sarbecoviruses – the viral subgenus containing SARS-CoV and SARS-CoV-2 – undergo frequent recombination and exhibit spatially structured genetic diversity on a regional scale in China. SARS-CoV-2 itself is not a recombinant of any sarbecoviruses detected to date, and its receptor-binding motif, important for specificity to human ACE2 receptors, appears to be an ancestral trait shared with bat viruses and not one acquired recently via recombination. To employ phylogenetic dating methods, recombinant regions of a 68 genome sarbecovirus alignment were removed with three independent methods. Bayesian evolutionary rate and divergence date estimates were shown to be consistent for these three approaches and for two different prior specifications of evolutionary rates based on HCoV-OC43 and MERS-CoV. Divergence dates between SARS-CoV-2 and the bat sarbecovirus reservoir were estimated as 1948 (95% highest posterior density (HPD): 1879–1999), 1969 (95% HPD: 1930–2000) and 1982 (95% HPD: 1948–2009), indicating that the lineage giving rise to SARS-CoV-2 has been circulating unnoticed in bats for decades.

Boni, M.F., Lemey, P., Jiang, X. *et al.* Evolutionary origins of the SARS-CoV-2 sarbecovirus lineage responsible for the COVID-19 pandemic. *Nat Microbiol* (2020). <https://doi.org/10.1038/s41564-020-0771-4> and <https://rdcu.be/b7nT0>