Review of: "Sequence evidence that the D614G clade of SARS-CoV-2 was already circulating in northern Italy in the fall of 2019"

Felix Broecker

Potential competing interests: No potential competing interests to declare.

The paper by Xuhua Xia discusses recent literature evidence that G614 mutants of SARS-CoV-2 were already circulating in Italy as early as October 2019 (around the same time as the first cases in Wuhan region occurred). The author suggests that therefore Italy is a possible origin of the COVID-19 pandemic. However, the conclusions are largely based on limited data from a single study that needs to be discussed more critically and put into context of other published literature. This would add value to the paper.

Specific comments:

Page 1, second to last sentence: "sequenced before the viral outbreak" is technically not true; the samples referred to in Table 1 were sequenced during the pandemic in 2021 (they were deposited in GenBank around August 2021) using archived samples originally used for measles and rubella surveillance. This is important since sequencing was performed using PCR amplicons which are very sensitive to environmental contaminations and may have amplified RNA contaminations from G614 strains circulating at that time.

The paper from which the sequences in Table 1 are derived is Amendola et al., 2022 (Molecular evidence for SARS-CoV-2 in samples collected from patients with morbilliform eruptions since late 2019 in Lombardy, northern Italy - ScienceDirect). Please add this citation to the reference list.

These few partial sequences that required nested PCR for amplification appear to be the basis for this paper’s claim that there were G614 mutants circulating in Italy in the fall of 2019. The limitations of the study Amendola et al., 2022 should be discussed.

Page 2: “The evidence from these SARS-CoV-2 sequence data from northern Italy is consistent with the dating of the MRCA to the summer of 2019.” Wasn’t this date derived mainly using sequence data from Chinese isolates, especially from those isolates close to the date of MRCA? In that case, the MRCA dated to the summer of 2019 is more likely reflecting the origin in China, not Italy.

Page 2: “It is also consistent with the twin-beginnings hypothesis on the origin, spread and evolution of SARS-CoV-2.” Is the hypothesis here that two highly similar coronaviruses coincidentally underwent animal-to-human transmission at around the same time in late 2019 in two distant parts of the world?

Page 3, bottom: “If scientists around the world would respond to WHO’s call to sequences archived samples (…)”. This is being done; here are some examples:

- PMID: 35899210 → seropositivity in people from Saudi Arabia who visited China before Nov 2019
- PMID: 34528756 → no evidence for SARS-CoV-2 infection in Valencia region, Spain, before March 2020
- PMID: 32620178 → no evidence for SARS-CoV-2 infection in California in Nov/Dec 2019
- PMID: 33207548 → no evidence for SARS-CoV-2 infection in Rome, Italy before March 2020

The findings of Amendola et al., 2022 should be discussed in the context of these papers.

Caption for Table 2 should read “D614G” not “G614G”

Figure 1 will make more sense if the most closely related coronaviruses from animals (such as bats and pangolins) are included. Are they any non-human coronaviruses that may be a potential common ancestor or that are more closely related to sequences in the lower part?