1   **Circular arguments on the origin of SARS-CoV-2**

2   David Bahry[*1]

3   [1] Department of Biology, Carleton University, 1125 Colonel By Dr, Ottawa, ON K1S 5B6, Canada
4   [*] email: davidbahry@cmail.carleton.ca

6   Two recent papers argue that SARS-CoV-2 emerged from wildlife sold in the Huanan Seafood

7   Market (HSM) with lineages A and B spilling over separately,[1,2] against alternatives such as a

8   research lab accident.[3] Yet they give circular arguments, by relying on data itself non-randomly

9   ascertained under local authorities' initial—understandable—assumption of a HSM spillover.

10   It remains plausible that HSM was not the origin of SARS-CoV-2, but only an early

11   super-spreader event for lineage B, "due to the high number of visitors every day."[4] Lineage A

12   shares its 'T/C' haplotype with outgroup cousins such as RaTG13, and so is plausibly ancestral,

13   whereas lineage B with its 'C/T' haplotype has dominated the pandemic including most of the

14   known early HSM-centered cases and HSM environmental samples, and emerging variants.[5]

15   Worobey et al. argue: "Early cases lived near to and centered on the Huanan market";

16   and within HSM, "Positive environmental samples [were] linked … to live mammal sales";

17   both claims dismiss ascertainment bias.[1] However, after only *four* reported HSM-linked cases,

18   Wuhan CDC initiated "a retrospective search for pneumonia patients potentially linked to the

19   market,"[6] including "epidemiology surveillance at several hospitals (close to Huanan market),

20   Huanan market and the neighborhood of Huanan market."[7] HSM sampling concentrated on

21   the south-western corner, where wildlife was sold, with most other stalls not sampled at all.[4]

22   On cases clustering around HSM, Worobey et al. claim: "We tested the robustness of

23   our results to the possibility of ascertainment bias."[1] However, their test only considered false

24   positives near the market, testing robustness to this possibility by dropping the cases nearest

25   the market from the data. Yet the issue was not false positives, but false negatives: cases missed

for *not* being near HSM. Their test is like surveying New Yorkers; dropping the 68% of New Yorkers nearest to Central Park; and concluding from the other 32% of New Yorkers that most of humanity lives near to Central Park, a result robust to the possibility of ascertainment bias.

On linking HSM environmental samples to live mammal sales, Worobey et al. claim that they "investigate the robustness of these findings to … over-sampling of live mammal and unknown meat stalls"; yet their Table S10 shows this only meant considering 2× sampling of these stalls. [1] This potentially far underestimates sampling bias, given the initial assumption of a HSM zoonosis. Without details of the sampling process—which Worobey et al. do not have—there is no basis for assuming so little focus on wildlife stalls, given early assumptions.

Worobey et al. also claim: "seroprevalence in Wuhan was highest in the districts around the market."[1] This misleads about their source, which found *lower* seroprevalence in Jianghan district, the HSM district, than in three other downtown districts [note: data for April 2020].[8]

Against the possibility that HSM was a super-spreader, which would be lineage-specific, Worobey et al. claim: "Both early lineages of SARS-CoV-2 were geographically associated with the market." Yet they base this on only three data points—two of them spurious. The single lineage A sample found in HSM (out of 1,380 samples, 73 positive[4]), after both lineages were already circulating in Wuhan, was likely introduced by a human. The man who stayed in a hotel near HSM was likely included in their source,[9] together with its nine HSM-linked cases, *because* the hotel was near HSM: more ascertainment bias. Nothing can be concluded from the remainder: one man with lineage A, living in downtown Wuhan, 2.31 km south of HSM.

Pekar et al. argue: "the inability to reconcile the molecular clock … with a lineage A ancestor without information from related sarbecoviruses (e.g., the recCA [recombinant common ancestor]) requires us to question the assumption that both lineages A and B resulted from a single introduction."[2] Yet it requires no such thing. Pekar et al. themselves consider the recCA-constrained root plausible, later using it to date the alleged two spillovers. We also

have a straightforward explanation for the molecular clock's rejection of a lineage A ancestor.

Molecular clock rooting by design assumes the root to be near the earliest-sampled genomes:

in this case, lineage B genomes sought and found because they were linked to or near HSM.[10]

    This comment is not intended to be exhaustive.

References

1. Worobey, M, Levy, JI, Serrano, LM, et al. The Huanan Seafood Wholesale Market in Wuhan was the early epicenter of the COVID-19 pandemic. *Science* 2022. https://doi.org/10.1126/science.abp8715

2. Pekar, JE, Magee, A, Parker, E, et al. The molecular epidemiology of multiple zoonotic origins of SARS-CoV-2. *Science* 2022. https://doi.org/10.1126/science.abp8337

3. Bloom, JD, Chan, YA, Baric, RS, et al. Investigate the origins of COVID-19. *Science* 2021; **372**: 694. https://doi.org/10.1126/science.abj0016

4. Gao, G, Liu, W, Liu, P, et al. Surveillance of SARS-CoV-2 in the environment and animal samples of the Huanan Seafood Market. *Research Square* 2022; published online Feb 25. https://doi.org/10.21203/rs.3.rs-1370392/v1 (preprint).

5. Rambaut, A, Holmes, EC, O'Toole, Á, et al. A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. *Nat. Microbiol.* 2020; **5**: 1403–1407. https://doi.org/10.1038/s41564-020-0770-5

6. The 2019-nCoV Outbreak Joint Field Epidemiology Investigation Team, Li Q. An outbreak of NCIP (2019-nCoV) infection in China — Wuhan, Hubei Province, 2019–2020. *China CDC Wkly.* 2020; **2**: 79–80. https://doi.org/10.46234/ccdcw2020.022

7. WHO. WHO-convened global study of origins of SARS-CoV-2: China Part: Annexes. Geneva: World Health Organization, 2021. Available from:

https://www.who.int/docs/default-source/coronaviruse/who-convened-global-study-of-origins-of-sars-cov-2-china-part-annexes.pdf.

8. He, Z, Ren, L, Yang, J, et al. Seroprevalence and humoral immune durability of anti-SARS-CoV-2 antibodies in Wuhan, China: a longitudinal, population-level, cross-sectional study. *Lancet* 2021; **397**: P1075–1084. https://doi.org/10.1016/S0140-6736(21)00238-5

9. Lu, R, Zhao, X, Niu, P, et al. Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *Lancet* 2021; **395**: P565–564. https://doi.org/10.1016/S0140-6736(20)30251-8

10. Bloom, JD. Recovery of deleted deep sequencing data sheds more light on the early Wuhan SARS-CoV-2 epidemic. *Mol. Biol. Evol.* 202; **38**: 5211–5224. https://doi.org/10.1093/molbev/msab246