



An Article in *Science* on Covid Origins Contains a Fundamental Error¹

Michael B. Weissman²

[LINK TO ABSTRACT](#)

The origin of the SARS-CoV-2 (SC2) virus is a crucial question because minimizing future pandemic risks requires knowing whether research-related risks have become comparable to zoonotic risks. In 2022, Jonathan E. Pekar and 28 other authors (Pekar et al. 2022) published an influential article addressing the origin of the SARS-CoV-2 (SC2) virus. The article—I’ll call it P2022—claims that Bayesian analysis of the RNA sequences found in early cases indicates that they came from two successful introductions from a non-human host rather than from one; they argued that the likelihood that two successful introductions to humans had occurred was higher than the likelihood that just one had occurred.

The purpose of this paper is not to treat the broader question of where SC2 originated. That topic requires multiple disparate lines of evidence. I narrowly address only the analysis of P2022, which used a Bayesian treatment of a specific model to infer “substantial support favoring separate introductions...BF = 4.3...” (where “BF” is short for “Bayes factor”) (Pekar et al. 2022). I show that conclusion to be based on a fixable mathematical error.

¹ I thank John Marden and Ellen Fireman for major editorial help with a related draft and Jamie Robins for early editorial suggestions. I am greatly indebted to Angus McCowan for noticing the imbalanced conditions and for helping throughout the process, especially by extracting the data on simulation sizes, a task beyond my capabilities. I hope that he will publish an analysis without the approximations that I had to use here to extract estimates from the published paper.

² Department of Physics, University of Illinois at Urbana-Champaign, Urbana, IL 61801.
mbw@illinois.edu. Orcid: 0000-0003-0081-4429

Although several coding implementation errors found by Angus McCowan have been corrected in a revised online version, a major error in Bayesian reasoning still appears twice in the remaining P2022 analysis (McCowan 2023; 2025). The essence of the error is that the likelihood of the favored two-introduction model was enhanced by excluding two of the detailed observed results used in evaluating the likelihood of the disfavored single-introduction hypothesis. I show that the error can be fixed using the data and model in P2022 with minimal extra ingredients. Fixing the error reverses the conclusion of the Bayesian analysis. That reversal somewhat reduces the plausibility of the wet market zoonotic account of the pandemic origin, often considered the principle alternative to a lab-leak origin. I will show that a piece of the correction large enough to invalidate P2022 is already evident in one of its figures. After fixing the errors, I discuss the significance of the widespread acceptance of a clearly erroneous result in a major paper on a major topic.

Context: Why does it matter?

P2022 has been an influential paper in the discourse about the origin of SC2. According to Google Scholar, it has been cited over 300 times. The work was featured in *The New York Times* (Zimmer and Mueller 2022) even before publication. After publication, P2022 would continue to form a key part of the arguments about Covid origins found in popular articles, including *Foreign Policy* (Rasmussen and Worobey 2022a) and *The Globe and Mail* (Rasmussen and Worobey 2022b). In part, that prominence has been based on a dubious rhetorical claim (Rasmussen and Worobey 2022a; Rasmussen and Worobey 2022b) that a double spillover is vanishingly rare for lab leaks.

P2022 was published as a companion paper to Worobey et al. (2022) (I'll call it W2022), a key paper presenting evidence of early superspreading from the Huanan Seafood Market (HSM), which was interpreted as evidence that SC2 spilled over to humans from some other animal at HSM. That SC2 spread to humans from some intermediate host at the HSM has been the most widely considered zoonotic account, although it is not the only possible one. In fact, an earlier paper (Pekar et al. 2021, 412) by some of the P2022 authors had suggested “this market cluster is unlikely to have denoted the beginning of the pandemic,” an assertion that was taken at that time to be consistent with other zoonotic sites. P2022 helped patch over difficulties for W2022's specific HSM spillover account, which I describe below.

One problem faced by the HSM account concerned the RNA sequences of the market-linked cases. These were all of lineage “B” rather than of the other early lineage “A”. These lineages differ only by two nucleotides, but in those two spots A

resembles a variety of related natural bat SARS-like coronavirus sequences (Pekar et al. 2022; Worobey et al. 2022). Since these two nucleotides are conserved in all the known relatives, the probability that B could be ancestral to all the human cases is low. For A to descend from B would require two improbable mutations reverting back to the ancestral sequence, unaccompanied by any of the more probable non-reversionary mutations (Caraballo-Ortiz et al. 2022; Bloom 2021; Lv et al. 2024; Bloom 2025; Pipes et al. 2020; Kumar et al. 2021). That early HSM cases were all of lineage B thus tended to support the widely-noted possibility (e.g. (Rouzioux 2025)) that HSM was not an introduction site but rather a superspreader site for a slightly mutated version that turned up downstream from an earlier introduction somewhere else of a version closer to the natural relatives (Caraballo-Ortiz et al. 2022; Bloom 2021; Lv et al. 2024; Bloom 2025). Introductions outside HSM could include either zoonotic or research-related ones.

P2022 address the issue of the HSM cases not seeming to be ancestral by arguing that Bayesian analysis of the sequence-based phylogeny indicates that B probably spilled over into humans *separately* from A. In principle, that would allow other types of evidence to show that B probably spilled over in HSM. That would still leave questions about the introduction leading to lineage A, but there is general agreement that if lineage B was introduced zoonotically at HSM then the origins were zoonotic regardless of the details of any other introductions.

Separate introductions of A and B would also mean that the most recent common ancestor (MRCA) had been in some previous host rather than in humans. That makes the estimated time of the introductions to humans later than it would be if both lineages came from one introduction. A later introduction time is more compatible with the HSM account than were previous estimates, such as the earlier one from the P2022 authors (Pekar et al. 2021). The P2022 claim that B probably descended from a separate introduction would thus ameliorate some difficulties for the specific zoonotic account in which HSM was an introduction site. It is important to note, however, that P2022 has little bearing on other possible zoonotic accounts. Because of this, my analysis showing that P2022 is incorrect should not be taken as proof that SC2 originated from a research-based setting.

The companion paper, W2022 (Worobey et al. 2022), sharing many coauthors with P2022, has also run into statistical trouble, although none as clear-cut as what I shall describe for P2022. Several researchers (e.g. (Bahry 2023)) have suggested that the case location data on which W2022 relied had been seriously skewed by ascertainment bias. I noticed that one statistic prominently featured in W2022 had the wrong sign for the W2022 model but the right sign for major ascertainment bias (Weissman 2024). Stoyan and Chiu (2024) found other statistical issues with W2022. In particular, rejection of an unrealistic null model of case locations was incorrectly taken by W2022 to imply that an HSM-based model was correct.

W2022 described the presence of SC2 RNA in swabs taken at HSM, but Bloom (2023) noted that the traces of SC2 RNA found in the swabs were negatively correlated with mtDNA of suspected non-human potential hosts. In contrast, the RNA levels from the four actual animal coronaviruses that were detected in appreciable quantities all correlated positively with their known hosts' mtDNA (Bloom 2023). Levin (2025) reanalyzed the various types of observations used in W2022 by more conventional Bayesian methods, claiming to find results strongly disfavoring the W2022 model and pointing to a source on the opposite side of the Yangtze River.

The original P2022 paper obtained a Bayes factor of about 60 favoring a two-successful-introduction model over a single-successful-introduction model. This result was described in the *Foreign Policy* article (Rasmussen and Worobey 2022a) as “a roughly 99 percent probability” that the two strands spilled over into humans separately. Three coding errors in those calculations were then noted on the “PubPeer” post-publication review site (McCowan 2023) by Angus McCowan (under a pseudonym) and later described by him in an arXiv paper (McCowan 2025). The current online version of P2022 now includes corrections for those three errors. The corrections reduced the likelihood ratio favoring two introductions from ~ 60 in the original version to a less compelling ~ 4.3 in the current version.

The most important remaining error, however, is not in coding but in basic Bayesian logic. This error was first mentioned by McCowan in February 2023 and elaborated on with specifics in March 2023 (McCowan 2023). McCowan's later arXiv paper used new modeling and simulations to correct this error, but it has not been corrected in the current P2022 appearing online in *Science*. The purpose of this paper is to show that the error can be explained and fixed without using new modeling or simulations. I use the model, data, and simulations of the original paper but correct the logic, leading to a reversal of the original conclusion. Thus, McCowan and I address and fix the same problem, but approach it differently.

In December of 2025, a brief “eletter” I wrote on this topic was published by *Science*. While the content of the eletter and this *Econ Journal Watch* article are similar, *Science* eletters are not peer-reviewed, do not receive “DOI” numbers, and are not indexed in major databases. Thus, they do not replace more formal publications. *Econ Journal Watch* employs peer review and encourages the authors of criticized papers to respond. I hope that by clarifying both the error and its fix this paper will encourage the editors of *Science* and others who had propagated the P2022 result to issue corrections.

The Central P2022 Argument

Figure 1, Figure S30 of P2022, illustrates an example of its central hypothesis. It is that SC2 had two successful spillovers to humans after circulating in some

intermediate host. One of those spillovers gave rise to the successfully propagating lineage A and the other to the successfully propagating lineage B. These two lineages differ by only two nucleotides out of ~30,000 in the entire sequence. P2022's claim is that its modeling supports "the hypothesis that lineages A and B represent separate introductions," which it denotes as hypothesis "I₂".

The competing hypothesis, I₁, is that both lineages came from a single successful spillover. The "Observed Phylogeny" box on the right of Figure 1 illustrates the I₁ hypothesis if its left-most sequence is taken to be the spillover sequence.

Figure 1: A pictorial guide to the type of scenarios used for the two-introduction hypothesis, reproduced from Figure S30 of P2022.

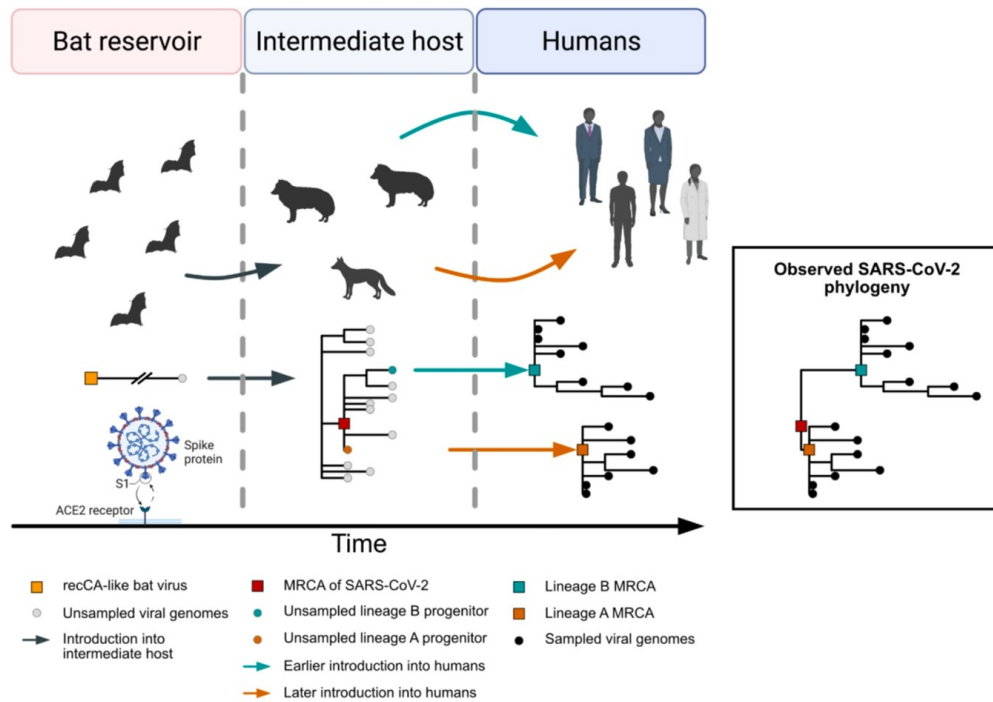


Figure S30. Schematic depicting the multiple zoonotic origin of SARS-CoV-2. A recCA-like virus was circulating in bats, and likely after gaining the ability to bind ACE2, jumped into an intermediate host. Therein, lineages A and B appeared and were separately introduced into humans shortly thereafter. An example phylogeny of viruses in the intermediate host is depicted, leading to separate phylogenies for lineages A and B. The resulting SARS-CoV-2 phylogeny from the combined lineage A and B viruses is presented in the black box. This scenario depicts a lineage A ancestral haplotype. See Figure S31 for intermediate and lineage B ancestral haplotypes.

Used under <https://creativecommons.org/licenses/by/4.0/>

P2022 presents a Bayesian analysis to compare the likelihoods of I_1 and I_2 . The standard Bayesian method to use evidence (e.g. the collection of RNA sequences found) is to compare the conditional probabilities of obtaining the observed results under the two hypotheses, i.e. compare the likelihoods of the hypotheses: $P(\text{data} | I_2)$ vs. $P(\text{data} | I_1)$. The odds $P(I_2)/P(I_1)$ are updated by the ratio $P(\text{data} | I_2)/P(\text{data} | I_1)$ for each new independent piece of data using Bayes' theorem.

$$\frac{P_{new}(I_2)}{P_{new}(I_1)} = \frac{P_{prior}(I_2) P(\text{data} | I_2)}{P_{prior}(I_1) P(\text{data} | I_1)} \quad (1)$$

It may be helpful to first illustrate the basic error with an analogy. If one were to update the odds for deciding which of two suspects committed a burglary using the observation that a blue Toyota had been seen leaving the scene it would not be correct to use the ratio $P(\text{drives car} | \text{suspect 2})/P(\text{drives blue Toyota} | \text{suspect 1})$. That comparison would incorrectly favor the “suspect 2” hypothesis since most cars are neither blue nor Toyotas. That fundamental error in logic is precisely analogous to the core error of P2022. In what follows I shall calculate the corrections to the I_2 likelihood used by P2022 when it is estimated using the same observed properties as were used for I_1 . That is, I shall use “drives blue Toyota” for both probabilities.

To make the likelihood comparisons, P2022 employs a stochastic model of how the virus transmits and mutates after a single introduction. Transmission from any case is highly over-dispersed in that some individuals have many possible transmission connections and others have few. Mutation is assumed to be a random Poisson process with about one mutation per four transmissions.

It is impractical to model the conditional probabilities of obtaining any specific large data set of sequences because those probabilities are far too low to be picked up in a reasonable number of simulations. Instead, P2022 uses simulations to calculate the likelihoods using some selected properties of the observed sequences.

Whenever a few properties of high-dimensional data are chosen post-hoc for statistical analysis, multiple-comparison issues arise that are not present for pre-specified analyses. Although that issue, shared by frequentist and Bayesian analyses, is relevant here, I shall dwell on a more unusual problem—that two of the features P2022 selected were only required for one of the hypotheses.

P2022 picks out two features of the observed phylogeny that do not seem to sit especially well with I_1 , the single-spill hypothesis. One is that the sequence set includes neither of the two possible intermediate sequences on the path between A and B, i.e. sequences differing from A and from B by one nucleotide each. One way that intermediates could be missing is that both single-nucleotide mutations

occurred in a single transmission step, which happens occasionally. Another way is that the 787 sequences available for P2022’s analysis just failed to pick up either intermediate sequences on the path from A to B or their intermediate descendants that may have been present in the tens of thousands of cases that occurred in the period (up to Feb. 14, 2020) during which the samples P2022 analyzed were taken. (More complete data did find intermediates (Lv et al. 2024), although they may not have been descended from ones on the path from A to B (Pekar et al. 2025).) The other feature is that two lineages have about the same size (i.e. number of cases). That might seem surprising if one lineage branched off from the other.

P2022 defines a “topology” of the sequences, which it denotes “ τ ,” to capture the size and sequence difference features and another property describing the relation of the lineages to their estimated MRCA. P2022 evaluates the probability of having this topology by looking at a set of “ascertained” samples randomly drawn from the first 50,000 sequences generated by each run of their stochastic model. The specific description from P2022 is:

a topology corresponding to a single introduction of an ancestral C/C haplotype—characterized by two clades, each comprising $\geq 30\%$ of the taxa, possessing a large polytomy at the base, and separated from the MRCA by one mutation was only observed in 0.0% of our simulations. Further, a topology corresponding to a single introduction of an ancestral lineage A or lineage B haplotype—characterized by a large basal polytomy and a large clade, comprising between 30 and 70% of taxa, two mutations from the root with no intermediate genomes—was observed in only 3.1% of our simulations (Pekar et al. 2022, 962)

To clarify, the main features chosen were that:

1. There are two lineages, with no intermediate lineages detected. Each lineage derives from a root sequence that constitutes “a large basal polytomy”, i.e. has at least a specified number of directly descendant branches from a single root. To illustrate, in Figure 1 (P2022’s S30) the sketch in the “Observed Phylogeny” box shows two small polytomies, one each at the roots of lineages A and B.
2. The numbers of cases in the two lineages are comparable. If we denote the ratio of the sizes of the two lineages as SR, P2022 requires $3/7 \leq SR \leq 7/3$ or $|\ln(\text{SR})| \leq \ln(7/3) = 0.85$.
3. The two lineages differ by 2 nucleotides, which I denote $D=2$.
4. The inferred MRCA is either one of the lineage roots or upstream of both lineage roots.

The central problem is that I_2 was not required to meet the second and third of these features required for I_1 , i.e. the conditions “comprising between 30 and 70% of taxa, two mutations from the root with no intermediate genomes.” Instead P2022 used only “ $P(\tau=(\tau_p, \tau_p) | I_n = I_2) = P(\tau=\tau_p | I_n=I_1)^2$ ”, i.e. condition (1), then adjusted later by a small factor to include condition (4), as evident in P2022 Table S5, discussed below.

Regardless of whether the P2022 model of post-introduction descent and random sampling is realistic, it is a specific, well-defined mathematical model applied to a specific data set. Therefore, it has well-defined implications. Like any mathematical implications, those can be calculated correctly or incorrectly.

Overview of the Remaining Corrections

It is essential that the same selected properties of the data be used for each hypothesis when Bayesian calculations update estimates of the odds of two hypotheses by taking the ratio of the conditional probabilities of those data for the two hypotheses, $P(\text{data} | I_1)/P(\text{data} | I_2)$. The conditional probability of more-detailed properties of the data (e.g. “blue Toyota”) will be smaller than the conditional probability of less-detailed properties (e.g. “car”).

P2022 calculated $P(\tau | I_1)$ from 1100 simulations of single introductions with its favored parameters. (These parameters were a nominal doubling time of 3.47 days, an ascertainment probability of 0.15, and 100 descendant branches used as the minimum for a polytomy.) In the current online version of P2022 only 34 simulations met the criteria used for τ , giving a point estimate $P(\tau | I_1) = 0.031$, with a 95% confidence interval of (0.022, 0.043).

P2022 did not model or simulate the two-introduction I_2 account, as was noted soon after publication (He and Dunn 2022). Instead, P2022 calculated $P(\tau | I_2)$ by using one of the properties of the I_1 simulations, whether an introduction produces a large enough basal polytomy. They denote that minimum-size basal polytomy criterion for a single introduction as “ τ_p ”. In the current version of P2022, 523 (47.5%) of the 1100 I_1 simulations meet that τ_p criterion. P2022 describes the process of extrapolating to $P(\tau | I_2)$ as follows:

We assume each introduction is independent, allowing us to generalize this probability to $P(\tau | I_n)$. For example, $P(\tau=\tau_p | I_n=I_1) = 0.475$ and $P(\tau=(\tau_p, \tau_p) | I_n = I_2) = P(\tau=\tau_p | I_n=I_1)^2 = 0.226$. (Pekar et al. 2022, supplement pg. 13)

This formula makes it unambiguously clear that the two introductions are meant to each consist of a single introduction of the type described by the simulations, since the numerical τ_p is directly taken from those single-introduction sim-

ulations. The likelihood ratio P2022 obtains is then proportional to $(P(\tau_p | I_1))^2 / P(\tau | I_1)$.

For each of nine different parameter sets used, P2022's Table S5 shows a Bayes factor of almost exactly $0.6 * (P(\tau_p | I_1))^2 / P(\tau | I_1)$. The factor of 0.6 was obtained from some constraints on the relation of the observed sequences to varying estimates of the probabilities of possible MRCA sequences, as mentioned in the fourth criterion listed above. To reiterate, the central problem with this method of estimating $P(\tau | I_2)$ is that it does not include the conditions “comprising between 30 and 70% of taxa, two mutations from the root with no intermediate genomes” required for I_1 (McCowan 2025; 2023). To be conservative and to keep the analysis simple, I shall ignore the complicated I_1 -favoring MRCA-dependent factor. Instead, I will just use the simple observed sequence difference $D=2$.

Consistently using the simple $D=2$ constraint rather than the more complicated and questionable constraint on the relation of the observed sequences to the hypothetical MRCA allows a major simplification in the analysis. As explained below, for I_2 meeting the size constraint depends only on the *difference* in the times of the two introductions but the mutation number constraint depends only on the *sum* of the two times after the MRCA to the two introductions. That decouples the two constraints, allowing each to be treated separately by simple methods.

Fixing the Size-Ratio Constraint

In the P2022 data the ratio of the number of sequences in the two lineages was $SR = 0.648/0.352 = 1.84$, with $\ln(SR) = 0.61$. For I_1 , P2022 required that a simulation produce a fraction of the sequences in the smaller lineage of at least 30%, i.e. $|\ln(SR)| \leq \ln(7/3) = 0.85$ in order to meet the topological criteria. In what follows I shall apply the same condition $|\ln(SR)| \leq 0.85$ to pairs of lineages consisting of pairs of P2022's simulated single introductions, just as P2022 used pairs of its single introductions to calculate the probability of a pair meeting the τ_p criterion.

The inequality condition P2022 used is reminiscent of the sorts of inequalities used in frequentist null hypothesis significance testing. Standard Bayesian likelihood calculations do not use such conditional probabilities of one-sided or mostly one-sided extensions of the observed results but rather ask what would be the probability or, for continuous variables, the probability density function (PDF) of the observed result for each hypothesis. That distinction turns out not to be important here. Both I_1 and I_2 give broad distributions of $\ln(SR)$, so $PDF(\ln(SR))$ is about constant over the range $0 \leq |\ln(SR)| \leq 0.85$. Thus, P2022's non-standard use of $P(3/7 \leq SR \leq 7/3 | I_1)$ gives a good approximation to use in the likelihood ratio. The use of p-value-like conditions rather than observed results can, however, lead to errors in other Bayesian applications.

Since the two introductions are described as independent, one can obtain a limit on the distribution of the ratio of the sizes of the two resulting lineages simply from the range of sizes generated by P2022's I_1 model. No new simulations or assumptions are needed. The narrowest distribution, and thus the highest probability of meeting the relative size condition, is obtained if neither lineage has a head start, i.e. the two introductions are simultaneous. The size distribution is broad enough for that simultaneity to also maximize the probability density near the observed SR.

The correction factor to $P(\tau|I_2)$ is then the fraction of pairs of simulation results for I_1 that have close enough sizes after a fixed time near the end of the simulation. Only the simulations that meet the τ_p criterion are relevant, since the others have already been excluded. P2022 does not make the size distribution of simulation outputs at some time easily available, but provides the information needed to get a good estimate in its supplementary Figure S22, shown below as Figure 2. It describes the distribution of times needed to reach various sizes. The key step in using Figure S22 is that simulations that take longer to reach a fixed size will be smaller after a fixed time.

The P2022 statistical tests use samples from the first 50,000 cases in each simulation. For a simulation pair to meet the size-ratio constraint, the larger one (50% to 70% of the total) should end up with about 30,000 cases. Figure S22D shows that once a simulation generates over 10,000 cases the growth becomes close to uniformly exponential with a doubling time a bit below the nominal value, 3.47 days. That increase in the doubling rate is to be expected because the cases gradually tend to become concentrated in the more-connected nodes. That exponential growth allows the distribution of times to reach a size near 30,000 to be converted to a distribution of logarithmic sizes at a fixed time by a simple conversion factor.

Here is an example: Say that one particular simulation happens to have taken the median length of time to reach $\sim 32,000$ cases, i.e. 15 doublings. According to Figure S22C, its cumulative doubling time, including the highly stochastic initial stages, is ~ 3.9 days, so those 15 doublings took $3.9 * 15 = 58.5$ days. A simulation at the 25th percentile has a cumulative doubling time of about 4.6 days, so it would take 69 days to reach that many cases. It is then 10.5 days behind. With a current (not cumulative) doubling time of 3.5 days it would be a factor of $2^{10.5/3.5} = 8$ short of the number of cases for the median simulation. Therefore, it would not be close enough in size for the pair to be counted as meeting the τ criteria. The simulation that grows at the median rate should have the best chance of having other simulations reach about the same size, but this example shows that substantially fewer than half of the other simulations are close enough in size to meet the relative size criterion. Just examining Fig S22 is enough to show that only a relatively small fraction of the pairs are close enough in size to meet the condition $|\ln(\text{SR})| \leq$

0.85.

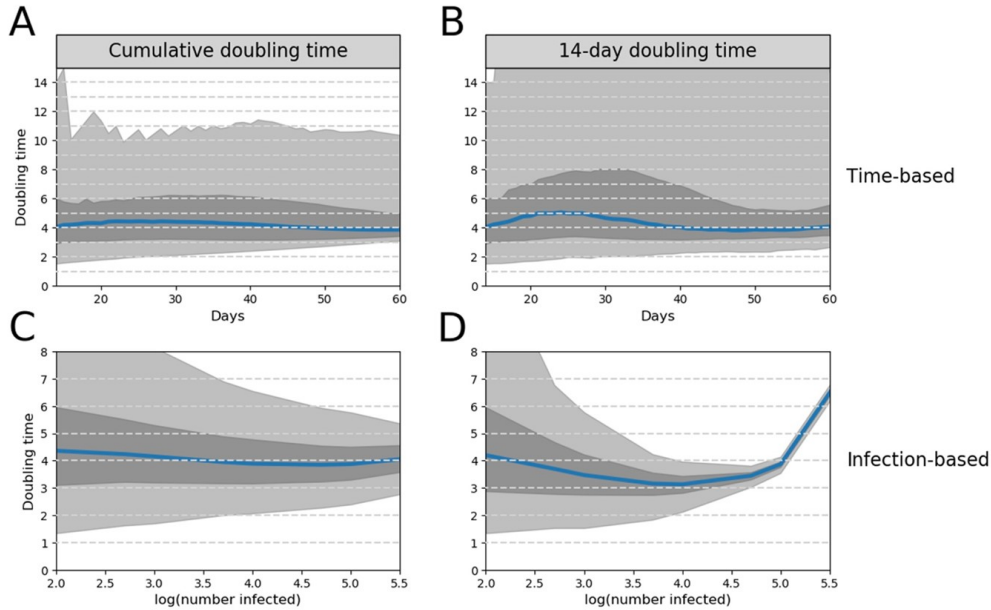
Figure 2: Figure S22 of P2022.

Figure S22. Inferred doubling times of simulated epidemics. Inferred doubling times of the 1100 primary simulations. (A) Cumulative doubling time since the start of the simulation. (B) 14-day doubling time from day 14 until the end of the simulation, with cumulative doubling time reported prior to day 14. (C) cumulative doubling time once a certain number of individuals are infected (*e.g.*, the cumulative doubling time at the 100th infection). (D) 14-day doubling time once a certain number of individuals are infected, with cumulative doubling time reported if that number of infections occurred before day 14 in the simulation. The center blue line represents the median doubling time across the simulations. Darker and lighter shading represent the 50% and 95% HDI, respectively.

Used under <https://creativecommons.org/licenses/by/4.0/>

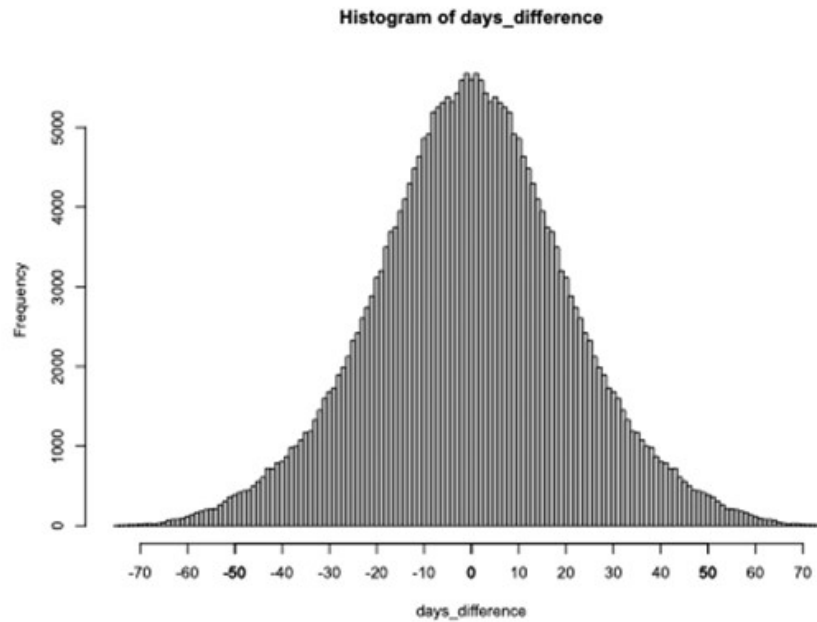
Now we can look at a more precise number. I will use the same essential technique, the nearly linear relation between the time to reach a large case number and the log of the number of cases at fixed times a little earlier. In order to meet the size-ratio constraint, the larger simulation has to produce less than 70% of the 50,000 cases used in the tests, so the relevant stopping point is about 35,000 cases. Jonathan Pekar has provided guidance on how to obtain the relevant files. McCowan has supplied a script (available in the Supplemental material to this paper) to extract the times to reach 35,000 cases for just those simulations that meet the τ_p constraint rather than for the larger set represented in P2022's Figure S22. (Simulations that do not meet the τ_p constraint cannot contribute to $P(\tau, I_2)$ regardless

of their sizes.)

One may get a feel for the distribution of times to reach 35,000 cases for the 523 simulations using the favored parameters that met the τ_p criterion from a histogram of the pairwise differences in those times, symmetrical about zero since the order of the simulations is arbitrary. Self-pairs are not included since they do not represent independent simulations.

A close look at Figure 22D shows that the current doubling time in the relevant range is ~ 3.25 days. The maximum allowed factor of $7/3$ in size then corresponds to a delay of $3.25 * \ln(7/3)/\ln(2) = 3.97$ days. Thus, simulation pairs with time differences greater than 4 days will not pass the size-ratio constraint. Of the $523 * 261 = 136,503$ non-self pairs, there are 24,818 pairs for which the completion times differ by four days or fewer. Nearly all of the 19,499 pairs with differences of 3 days or fewer will be close enough in size, but only approximately half of the 5319 pairs with nominal differences of 4 days will. The fraction of pairs with close enough size will then be ~ 0.162 .

Figure 3: The histogram of differences of the times to reach 35,000 cases for pairs of P2022's 523 simulation runs that met the τ_p criterion.



It is instructive to take a closer look at the information in Figure S22 to see what a careful reader could infer without bothering to access the P2022 supplemental files. As I show in my Supplement 1, simple inspection of Figure S22 reveals that $\sim 15\%$ of all the simulation pairs will be close enough in size. Since the set of simulations that meet τ_p is concentrated in the larger half of the size distribution, the relevant fraction could be almost twice that, with a mid-point estimate being $\sim 20\%$. Something like that estimate should be evident to a careful reader.

Applying the correction factor 0.162 directly to the P2022 Bayes factor would reduce it from 4.3 to 0.70, eliminating any tendency for the analysis to favor I_2 . That value, 0.70, includes both the polytomy condition and the MRCA-related condition. It corresponds to requiring I_2 to meet conditions (1), (2), and (4) of our list of the four conditions that were required for I_1 .

Now it is time to consider condition (3), the sequence difference. Since this condition partially overlaps condition (4) related to the MRCA, going forward I shall drop that I_1 -favoring MRCA condition. The fraction of simulation pairs in which each simulation met the τ_p criterion was 0.226, so including the relative size criterion lowers that by a factor of 0.162 to 0.0366. Thus, only 3.66% of the simulation pairs meet both the τ_p and size ratio conditions, (1) and (2) from our list of four conditions. The overall Bayes factor from including just those two conditions for I_2 (dropping the MRCA condition) would be $0.366/0.031 = 1.18$.

Fixing the Sequence-Difference Constraint

P2022 requires that the I_1 simulations produce a sequence difference of $D=2$ between the lineages in order to meet the τ criteria. That same condition needs to be applied to I_2 .

Unlike for the relative size condition, the distinction between this proper Bayesian condition of matching the observed $D=2$ and an improper one-sided p-value-like extension of the condition is crucial. I_1 tends to give small D since the biggest branches need to start early, before many mutations have occurred. I_2 can give arbitrarily large D because there are no relevant limits on the pre-introduction sequence diversity. It is necessary to use the actual result, $D=2$, not an arbitrary extension, since a one-sided extension of the observed $D=2$ to $D < 3$ would give a completely different likelihood ratio than would an extension to $D > 1$.

For I_2 , as for I_1 , the observed $D=2$ is not especially easy to meet. Visual inspection of P2022's Figure S30 gives a feel for why. Two sequences happen to spill over from the hypothetical host to humans. There are 28 possible introductory pairs derived from the MRCA including self-pairs, which there is no reason to exclude. Taking the sketch to represent sequence differences by horizontal lengths, we observe that only 6 of the 28 pairs (21%) have $D=2$.

We need a quantitative analysis of $P(D=2|I_2)$ rather than counts from a

sketch. Each introduction is of a sequence descended from the MRCA, whatever that might happen to be. There are a very large number of possible mutations ($\sim 90,000$) that could have occurred, but one finds only the small observed value $D=2$ between the two lineages. Although the probabilities of individual mutations are far from equal, P2022 shows hundreds of mutations detected, none of which are at all likely to occur in any single line of descent on the time scale of weeks. Thus, the probability for each mutation on a path from the MRCA to an introduction must be quite small. There is no indication that the two mutations separating A and B are strongly linked to each other, since P2022's molecular clock model for mutations in humans includes no correlations between them, and because intermediate sequences appeared soon in real cases (Lv et al. 2024).

The sequence difference D_0 between I_2 's two introductions thus is the number of occurrences of very low probability events drawn from a large pool of possibilities. Therefore, the probabilities $P(D_0)$ should be very close to the Poisson limiting form. $P(D_0)$ is then fully characterized by its expectation value $E(D_0)$. $E(D_0)$ depends on the sum of the times from the MRCA to the two introductions multiplied by the mutation rate in the hypothetical prior host, but we do not need to consider those factors separately.

Mutations between the introduction and the detected lineage root can also contribute to the net D . These, like D_0 , will have a distribution of values around their mean and thus cannot be fine-tuned to produce $D=2$. For our purposes here, the Poisson distribution approximation for D should suffice. In the Supplement, I present a more complete analysis including the post-introduction pre-clade-root mutations. Although the effects of including that extra step are negligible for I_2 , they could be important for more complicated hypotheses involving more than two successful introductions.

The maximum possible Poisson $P(D=2|I_2)$ is found for $E(D)=2$, giving $P(D=2|I_2) \leq 2/e^2 = 0.271$. That is less than half of the MRCA-related factor of ~ 0.6 used in the complicated P2022 analysis.

Reaching that maximum Poisson probability requires fine-tuning $E(D)$ to precisely 2, without any prior justification. If instead we consider a plausible prior distribution for $E(D)$, call it $\rho_0(E(D))$, we may obtain a posterior distribution $\rho(E(D))$ by conditioning on the observed $D=2$ and then calculating $P(D=2)$ by marginalizing over $\rho(E(D))$. In effect, this procedure still allows some post-hoc adjustment of the I_2 model to agree with the observations, but it does not allow unrealistic fine-tuning. The procedure, (see Supplement 3), gives $P(D=2|I_2) = 0.193$.

Other Statistical Issues

McCowan (2023) has discussed further minor coding fixes required to bring the code into alignment with the algorithm described in the P2022 text, as well as

refinements of the precision obtained by using 110,000 simulations. Those steps can result in further small changes in $P(\tau | I_2)/P(\tau | I_1)$.

The main estimate used in P2022 is based on sets of 7,500 sequences drawn from simulated sets of 50,000. The actual data used, however, have only 787 sequences. Use of more realistic sparse sampling might also change $P(\tau | I_2)/P(\tau | I_1)$, particularly since the especially low rate of detection in early cases (Tsang et al. 2020) and the non-random sampling (Bloom 2025) might increase the probability of missing early intermediates. Since the entire two-spill hypothesis rested on the lack of detected early intermediates in humans, this correction for sparse non-random sampling could be important.

Proposed Modifications of the P2022 Hypothesis

Hensel (2025) has responded to my eletter in *Science* with some proposals for how some version of a picture with multiple successful introductions might end up being probable. Although these proposals are not directly relevant to my description of the P2022 errors, in Supplement 5 I discuss them for possible future reference.

Statistical Bottom Line

Incorporating the size-ratio and sequence-difference constraints for I_2 using parameters chosen to favor I_2 gives a corrected $P(\tau | I_2) \leq 0.475^2 * 0.162 * 0.193 \leq 0.0071$. The approximate resulting likelihood ratio is at most $0.0071/0.031 = 0.23$. In other words, the likelihood ratio is at least $0.031/0.0071 = 4.4$ favoring I_1 over I_2 .

Correcting P2022's major error in basic Bayesian techniques reverses the direction of the conclusion. Based on the P2022 model and data, a single introduction is more likely, as Lv et al. (Lv et al. 2024) concluded from their interpretation of more complete data.

Informal Cultural Reflections

That 29 authors, the editors of *Science*, and the authors of over 300 works citing P2022 overlooked such a major flaw in the paper raises several general issues. For one, our scientific error-correction process is not working well. Although the absence of a model for the two-spill hypothesis was noted soon after P2022 was published (He and Dunn 2022), no one (including me) pinpointed the plain-sight logic error until McCowan, a patent expert with no academic affiliations, pointed it out long after the paper had been accepted and widely publicized. Even though most of us tend to look over the abstract of a paper and skim through to find

some key points, it is still surprising that no academic caught the unambiguous logical problem with P2022. Perhaps a generally low level of statistical literacy contributed to our collective failure. The history of this prominent paper may provide a case study to remind authors, editors, readers, and educators that the availability of sophisticated software does not obviate the need for attention to basic statistical reasoning.

Historically, when a paper had a clear error that did not require innovative research to find and fix, journals had a technical comment category for corrective responses. The journal *Science*, along with all publications of the American Association for the Advancement of Science (AAAS), eliminated its “Technical Comment” paper category several years ago, leaving no standard procedure in those journals to correct even basic errors in mathematics or logic. This AAAS policy needs to be reversed.

An additional motivation makes error-correction particularly difficult for work related to Covid origins. Covid has already caused well over 20 million deaths, enormous economic disruption, and widespread ongoing adverse health effects. Covid may have leaked from some risky research project (In my opinion (Weissman 2025a) that is the most likely origin). Unsurprisingly, discussions of this possibility among scientists have been subject to unusual pressures. Unfortunately, the question of whether research may have led to the pandemic has been routinely framed as one of defending science as a whole against outside attacks, e.g. in the *Foreign Policy* article (Rasmussen and Worobey 2022a). That move is understandable since the scientific community has in fact been under attack, largely because some genuine scientific results have inconvenient implications, especially concerning the dangers of anthropogenic global warming. The problem, however, with a reflexively defensive approach is that when a scientific community abandons error-correction, that community is no longer doing science.

Supplements

Supplement 1

This Supplement uses Figure S22C to estimate the size-ratio correction.

P2022’s Figure S22C shows the width of the time distribution near the end of the simulations, e.g. at $10^{4.5}$ sequences, corresponding to 15 doubling times. The median cumulative doubling time at that point, including all the early more stochastic steps, is 3.9 days. Figure S22C shows the intervals containing 50% and 95% of the cumulative doubling times for the collection of simulations, giving about (3.2, 4.7) and (2.2, 6.0) days, respectively. These would be consistent with Gaussian approximations with standard deviations of 1.13 days and 0.97 days, respectively. The

form of the tails of the distribution will have negligible effects on the fraction of pairs that are close enough in size. For our approximate purposes here a Gaussian distribution with width 1.05 days should be adequate.

Thus we may approximate the distribution $\rho(\ln(\text{size}))$ toward the end of the simulation by using a Gaussian with standard deviation $15 \times 1.05 \text{ days} \times \ln(2) / 3.47 \text{ days} = 3.15$. This estimate may understate the width of the size distribution since Figure S22D shows that in the relevant window of case numbers the doubling time is roughly 3.2 days rather than the nominal 3.47 days.

The distribution then has the approximate form:

$$\rho(\ln(\text{size})) = \left(\frac{1}{3.15} \right) (2\pi)^{-0.5} e^{-\left(\frac{0.5}{9.90}\right)(\ln(\text{size}))^2} \quad (2)$$

If all the I_1 simulations contributed to the results that met the other criteria, then the distribution of $\ln(\text{SR})$ would just be a Gaussian with width $3.15 \times 2^{0.5} = 4.45$ since the sizes of the lineages descended from the two introductions would vary independently. Only 15% would then have $|\ln(\text{SR})| \leq \ln(7/3) = 0.85$, i.e. have $|z| \leq 0.85/4.45$ on a normal distribution.

The subset of simulations meeting the requirement of having a minimum polytomy size will tend to have a narrower distribution of sizes since the number of taxa and the basal polytomy's number of branches are likely to be correlated. One may make an extreme allowance for such an effect by eliminating the smaller half of the size distribution. Taking the integral numerically, 28% of the remaining pairs would then fall within the required size-ratio range. A cautious estimate of the correction for the lineage-size size constraint would then be a factor in the range 0.15 to 0.28, with a geometric mean of 0.20.

Supplement 2

This Supplement includes the Python script for extracting time-to-completion data, the list of relevant times, and the R program for finding how many were close enough.

The Python script gives the file of the days to get to 35k cases for simulations that meet the polytomy requirement, using “101” for those that do not reach 35k in 100 days.

```
#!/usr/bin/env python3
# coding: utf-8
# ##### Identify the runs with basal polytomies
# Download the clade analysis results from https://github.com/sars-cov-2-origins/multi-introduction/raw/refs/heads/main/FAVITES-COVIDLite/cumulative_results/FAVITES_results.zip.
# Unzip in the same directory as this script.
import os
runs_with_polytomies = []
```

```

for i in range(1, 1101):
    fpath = os.path.join("clade_analyses_CC", f"{i:04d}_clade_analysis_CC_polytomy.txt")
    with open(fpath) as f:
        n = sum(1 for _ in f) # count lines in the file
        if n >= 100: # each line represents a subclade or basal lineage
            runs_with_polytomies.append(i) # record if 100 branches or more
print(f"Number of runs with basal polytomies: {len(runs_with_polytomies)}")
print(f"Frequency of basal polytomies: {100*len(runs_with_polytomies)/1100:.1f}%")

# #### Get day of 35kth infection for the runs with basal polytomies
# Download the simulation results from https://github.com/sars-cov-2-origins/multi-
# introduction/raw/refs/heads/main/FAVITES-COVIDLite/cumulative_results/FAVITES_GEMF_dict.pickle.zip.
# Unzip in the same directory as this script.
import pickle
with open("FAVITES_GEMF_dict.pickle", "rb") as f:
    gemf = pickle.load(f)
results = {}
for run in runs_with_polytomies:
    run_id = f"{run:04d}"
    results[run_id] = 101 # default to end of simulation
    for day in range(101): # 0..100 days of simulation
        if gemf[run_id][day]["S"] < 4965000:
            results[run_id] = day
            break
#### write as single line
with open("day_of_35k_single_line.csv", "w", newline="") as f:
    writer = csv.writer(f)
    writer.writerow([results[k] for k in results])

```

Here's the R program and the list of completion dates.

```

> days<-c(37,51,63,65,49,42,45,58,64,73,71,55,36,38,62,75,47,46,46,74,35,71,51,52,55,54,59,
87,31,43,49,49,61,85,38,66,46,61,39,54,32,43,47,54,43,46,46,43,42,42,42,32,36,68,59,49,61,
64,59,46,36,58,80,35,71,41,66,47,43,88,58,42,49,48,36,84,53,53,59,49,52,49,61,57,77,53,37,
49,47,67,69,95,39,33,52,89,34,43,41,80,52,51,61,71,53,51,61,55,79,56,44,44,58,73,58,66,49,
56,90,50,61,62,53,69,67,48,51,55,43,68,46,35,61,47,48,40,70,77,60,52,38,56,92,51,78,87,79,
87,51,58,46,68,50,62,48,38,63,52,71,63,58,46,49,56,49,45,89,62,50,36,53,99,64,43,71,59,38,
33,72,47,57,31,34,29,54,63,41,71,49,59,35,26,55,56,73,59,44,43,43,68,50,48,43,42,90,41,
57,48,74,58,84,60,53,60,61,43,100,44,48,47,48,57,40,37,64,74,59,44,40,40,62,50,54,28,45,54,
41,42,57,46,38,57,52,30,42,78,50,49,36,48,51,41,44,32,59,37,51,66,67,91,53,61,60,51,53,57,
59,59,37,45,56,40,63,34,52,39,35,81,55,57,57,94,62,67,48,47,39,61,49,44,51,49,36,46,39,54,
63,48,60,73,54,59,65,77,75,63,60,49,48,32,41,74,56,79,56,67,51,58,42,59,43,49,44,41,51,64,
40,54,51,75,28,72,54,54,91,69,34,43,37,64,57,57,36,54,48,75,51,51,60,43,66,54,73,55,50,59,
86,49,47,45,40,50,53,50,82,32,75,68,51,47,93,63,62,52,39,42,77,50,53,44,45,64,45,100,70,48,
52,101,54,66,70,37,31,67,47,60,51,46,48,73,66,62,40,31,79,74,60,44,44,48,44,48,40,66,43,58,
41,44,69,37,53,49,68,49,47,60,66,57,59,60,39,67,30,67,92,59,55,49,44,36,49,60,49,59,46,45,
49,38,53,53,42,50,49,59,36,27,42,55,37,37,68,50,49,51,57,92,74,76,81,32,31,94,57,35,63,56,
80,41,30,52,39,29,50,41,42,42,50,87,45,69,41,32,72,76,28,51,41,57,51,63,92,81,59,50,43,31,
70,65,58,52,67,55,66,84,43,42,46,36,52,50,71,74)
> L<- length(days)
> count <- 0
> for (i in 1:(L-1)) {{for(j in (i+1):L) {if (abs(days[i]-days[j]) <= 3) {count <- count + 1}}}}
>count
[1] 19499
>for (i in 1:(L-1)) {{for(j in (i+1):L) {if (abs(days[i]-days[j]) == 4) {count <- count + 0.5}}}}
> count
[1] 22158.5

```

```
> 2*count/(L*(L-1))
[1] 0.1623298
```

Several effects can lead to minor changes in the factor of 0.162. The random draws of 7500 sequences from each 50,000 generated will slightly increase the variance of $\ln(\text{SR})$. Omission of the variance picked up in the last factor of 2 or so growth by the smaller lineage will slightly reduce the variance of $\ln(\text{SR})$.

Supplement 3

This Supplement describes marginalizing over plausible priors for $E(D)$ to obtain $P(D=2|I_2)$, assuming a Poisson distribution for D .

For brevity, I shall denote $E(D) = x$. One plausible prior $\rho_0(x)$ would be simply uniform up to truncation at some irrelevant large value. One then obtains $\rho(x) = x^2 e^{-x}/2$. Marginalizing over that gives $P(D=2) = 3/16$, close to the accidental value found from Figure S30. This factor is a bit smaller than the fine-tuned $2/e^2$. Another plausible prior would be uniform in $\ln(x)$, again truncated at irrelevant extreme values. It gives $\rho(x) = x e^{-x}$. Coincidentally, marginalizing over that also gives $P(D=2) = 3/16$.

For a general power-law prior $\rho_0(x)$ proportional to $x^{-\alpha}$ with α in the range $\alpha < 3$ for which divergences are not problematic $P(D=2) = (4-\alpha)(3-\alpha)/2^{(6-\alpha)}$. The maximum is obtained for $\alpha = (7*\ln(2)-2-(4+(\ln(2))^2)^{0.5})/(2*\ln(2)) = 0.53$, which gives $P(D=2) = 0.193$

Supplement 4

This Supplement describes how including the mutations that occur on the path from the introduction to the detected clade root changes $P(D=2|I_2)$. This extra mutation number is not necessarily describable by Poisson statistics because the clade roots are not randomly selected but are chosen by the Pekar algorithm in a way that depends on their difference from other sequences.

The resulting change in $P(D=2|I_2)$ turns out to be quite small. Inclusion of these post-introduction pre-root mutations may be more important in evaluating probabilities of more complicated multi-introduction hypotheses since this post-introduction diversity is inherent in the P2022 single-spill model and cannot be altered by adding more complications to the hypothesized pre-introduction stage.

We can denote the probabilities of the detected root being (0,1,2) nucleotides different from its introductory sequence by (p_0, p_1, p_2) . To obtain $D=2$ one needs that the sum of the Poisson pre-introductory D_0 and the two post-introductory changes be two. Adding the probabilities for each such path, again calling $E(D_0) = x$, gives:

$$P(D=2|I_2) = e^{-x} (2 p_0 p_2 + p_1^2 + 2 p_0 p_1 x + 0.5 p_0^2 x^2).$$

The maximum is obtained when x is fine-tuned to $x = (p_0 - 2 p_1 + (p_0^2 + 2 p_1^2 - 4 p_0 p_2)^{0.5}) / p_0$.

McCowan has extracted the post-introduction diversity distribution from the 1100 relevant P2022 simulations. The simulations give $(p_0, p_1, p_2) = (0.589, 0.235, 0.085)$ for the probabilities of 0, 1, or 2 mutations, the most important values. Then the fine-tuned maximum $P(D=2|I_2)$ occurs for $x=1.06$ which gives $P(D=2|I_2) = 0.22$, less than the 0.27 obtained for the fine-tuned maximum when the post-introduction diversity was ignored. Considering power-law priors for x , just as we did without including the post-introduction diversity, reduces the maximum to 0.176, slightly smaller than the value obtained before. Thus inclusion of the post-introduction diversity makes little difference for $P(D=2|I_2)$.

Recently, as the extreme problems with the P2022 analysis have become clearer, suggestions have arisen for more complicated stories of multiple successful introductions (Hensel 2025). Multiple introductions of a single sequence can raise the probability that the lineage with that sequence will end up with a size not too far below the median. For each such indistinguishable introduction, however, there is a probability $1 - (p_0^2 + 2 p_0 p_1) = \sim 0.38$ of getting an additional distinguishable clade root 2 nt or more different from the others in the P2022 model. Any attempt to restore the probability of multiple successful introductions by replacing the P2022 approach with a hypothesis of more than two successful introductions would need to account not only for the extra adjustable parameters required but also for the effects of such extra distinguishable clades on the topology.

Supplement 5

This Supplement addresses the recent eletter by Hensel (2025).

One of Hensel's suggestions concerns the $D=2$ condition, for which he proposes that the definition of I_2 should already include that it could not produce $D=0$ or $D=1$. That step would be inconsistent with any sort of Bayesian hypothesis evaluation, in which the likelihood of each hypothesis depends on the fraction of its outputs that match the observation. Excluding by fiat outputs that do not match the observations would allow any hypothesis, including I_1 , to reach a likelihood of 1.0.

Hensel's other two suggestions involve replacing P2022's I_2 hypothesis with more complicated multi-introduction accounts that contradict the explicit P2022 I_2 description and calculations based on pairs of independent single introductions. Although those proposals are not directly relevant to the central point of this paper some discussion of them may be useful in the future.

One suggestion was that, unlike in the P2022 description and calculations, the two introductions were not independent but instead had systematically similar

early-stage growth rates because they both arose in the same market. That *ad hoc* fix of the size ratio factor fits very poorly with the observation that was the original impetus for P2022—that 12 lineage B cases were detected in HSM in contrast to zero lineage A cases.

Hensel's other suggestion is that, unlike the P2022 model in which the probability of each lineage meeting the polytomy constraint was taken directly from the single-introduction simulations, there may have been multiple successful introductions of each lineage. That would tend to increase the likelihood of meeting the polytomy condition and might tend to reduce the spread in the size ratios. Such a Ptolemaic model not only involves adding multiple adjustable parameters but also raises new ways in which the multiple introductions could produce a topology different from that observed. I discuss that issue briefly in Supplement 4.

References

- Bahry, David.** 2023. Rational Discourse on Virology and Pandemics. *mBio* 14(3): e00313-23. [Link](#)
- Bloom, Jesse D.** 2021. Recovery of Deleted Deep Sequencing Data Sheds More Light on the Early Wuhan SARS-CoV-2 Epidemic. *Molecular Biology and Evolution* 38(12): 5211–5224. [Link](#)
- Bloom, Jesse D.** 2023. Importance of quantifying the number of viral reads in metagenomic sequencing of environmental samples from the Huanan Seafood Market. *Virus Evolution* 10(1): vead089. [Link](#)
- Bloom, Jesse D.** 2025. The Data are Insufficient to Confidently Root the SARS-CoV-2 Phylogenetic Tree. *Molecular Biology and Evolution* 42(6): msaf118. [Link](#)
- Caraballo-Ortiz, Marcos A., Sayaka Miura, Maxwell Sanderford, Tenzin Dolker, et al.** 2022. TopHap: rapid inference of key phylogenetic structures from common haplotypes in large genome collections with limited diversity. *Bioinformatics* 38(10): 2719–2726. [Link](#)
- He, Mai, and Lucia Dunn.** 2022. Assessing Probabilities for the Two Leading SARS-CoV-2 Origin Narratives. eletter to, “The molecular epidemiology of multiple zoonotic origins of SARS-CoV-2,” Jonathan Pekar et al., *Science* 377. [Link](#)
- Hensel, Zach.** 2025. Clarifying the Hypotheses Tested in Pekar et al. (2022). eletter to, “The molecular epidemiology of multiple zoonotic origins of SARS-CoV-2,” Jonathan Pekar et al., *Science* 377. [Link](#)
- Kumar, Sudhir, Qiqing Tao, Steven Weaver, Maxwell Sanderford, et al.** 2021. An Evolutionary Portrait of the Progenitor SARS-CoV-2 and Its Dominant Offshoots in COVID-19 Pandemic. *Molecular Biology and Evolution* 38(8):

3046–3059. [Link](#)

- Levin, Andrew T.** 2025. A Bayesian Assessment of the Origins of COVID-19 using Spatiotemporal and Zoonotic Data. *NBER Working Paper* 33428. National Bureau of Economic Research (Cambridge, MA). [Link](#)
- Ly, Jia-Xin, Xiang Liu, Yuan-Yuan Pei, Zhi-Gang Song, et al.** 2024. Evolutionary trajectory of diverse SARS-CoV-2 variants at the beginning of COVID-19 outbreak. *Virus Evolution* 10(1): veae020. [Link](#)
- McCowan, Angus.** 2023. comment to “The molecular epidemiology of multiple zoonotic origins of SARS-CoV-2,” Jonathan E. Pekar et al., *Science* 377(6609). [Link](#)
- McCowan, Angus.** 2025. Purported quantitative support for multiple introductions of SARS-CoV-2 into humans is an artefact of an imbalanced hypothesis testing framework. *arXiv* 2502.20076. [Link](#)
- Pekar, Jonathan E., Niema Moshiri, Philippe Lemey, Alexander Crits-Christoph, et al.** 2025. Recently reported SARS-CoV-2 genomes suggested to be intermediate between the two early main lineages are instead likely derived. *Virus Evolution* 11(1): veaf008. [Link](#)
- Pekar, Jonathan E., Andrew Magee, Edyth Parker, Niema Moshiri, et al.** 2022. The molecular epidemiology of multiple zoonotic origins of SARS-CoV-2. *Science* 377(6609): 960–966. [Link](#)
- Pekar, Jonathan, Michael Worobey, Niema Moshiri, Konrad Scheffler, and Joel O. Wertheim.** 2021. Timing the SARS-CoV-2 index case in Hubei province. *Science* 372(6540): 412–417. [Link](#)
- Pipes, Lenore, Hongru Wang, John P Huelsenbeck, and Rasmus Nielsen.** 2020. Assessing Uncertainty in the Rooting of the SARS-CoV-2 Phylogeny. *Molecular Biology and Evolution* 38(4): 1537–1543. [Link](#)
- Rasmussen, A., and M. Worobey.** 2022a. Conspiracy Theories About COVID-19 Help Nobody. *Foreign Policy*, September 15. [Link](#)
- Rasmussen, Angela, and Michael Worobey.** 2022b. COVID-19 almost certainly did not come from a lab leak. Here’s how we know. *The Globe and Mail*, July 28. [Link](#)
- Rouzioux, C.** 2025. De l’Origine du SARS-CoV-2 aux risques de zoonoses et de manipulations dangereuses de virus. *Bulletin de l’Académie Nationale de Médecine* 209(5): 650–662. [Link](#)
- Stoyan, Dietrich, and Sung Nok Chiu.** 2024. Statistics did not prove that the Huanan Seafood Wholesale Market was the early epicentre of the COVID-19 pandemic. *Journal of the Royal Statistical Society Series A: Statistics in Society* 187(3): 710–719. [Link](#)
- Tsang, Tim K., Peng Wu, Yun Lin, Eric H. Y. Lau, et al.** 2020. Effect of changing case definitions for COVID-19 on the epidemic curve and transmission

parameters in mainland China: a modelling study. *The Lancet Public Health* 5(5): e289–e296. [Link](#)

Weissman, M. B. 2024. Proximity ascertainment bias in early COVID case locations. *Journal of the Royal Statistical Society Series A: Statistics in Society* 187(3): 720–722. [Link](#)

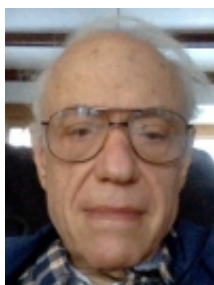
Weissman, M. B. 2025a. An Inconvenient Probability. *Michael's Substack*, Substack, March 14. [Link](#)

Weissman, M. B. 2025b. Balancing the conditions used in the Bayes factor for the spillover number. eletter to, “The molecular epidemiology of multiple zoonotic origins of SARS-CoV-2,” Jonathan Pekar et al., *Science* 377. [Link](#)

Worobey, Michael, Joshua I. Levy, Lorena Malpica Serrano, Alexander Crits-Christoph, et al. 2022. The Huanan Seafood Wholesale Market in Wuhan was the early epicenter of the COVID-19 pandemic. *Science* 377(6609): 951–959. [Link](#)

Zimmer, Carl, and Benjamin Mueller. 2022. New Research Points to Wuhan Market as Pandemic Origin. *New York Times*, February 26. [Link](#)

About the Author



Michael B. Weissman is a Professor Emeritus of Physics at the University of Illinois at Urbana-Champaign and a Fellow of the American Physical Society. He has worked mostly on condensed matter experiments, especially on low-frequency noise in disordered materials. He has worked in botany and chemistry departments. His undergrad degree is in math. Since retirement he has used statistical forensics to help catch a couple of fake pollsters and has written papers on numerous errors in causal reasoning in physics education research publications. He did time for Vietnam draft resistance and was an originator of the scientists’ boycott of the Strategic Defense Initiative. He thinks that libertarian philosophy is factually detached from the realities of our species, implicitly cruel, and will lead to environmental catastrophe. He supports unions and is considering changing his opposition to the designated-hitter rule. That change may be the result of neurological sequelae to Covid. His email is mbw@illinois.edu.

[Go to archive of Comments section](#)
[Go to March 2026 issue](#)