










Waiting for the truth: is reluctance in accepting an early origin hypothesis for SARS-CoV-2 delaying our understanding of viral emergence?

Marta Canuti ¹, Silvia Bianchi ^{1,2}, Otto Kolbl,³
Sergei L Kosakovsky Pond ^{4,5}, Sudhir Kumar ^{4,5,6}, Maria Gori ^{1,2},
Clara Fappani ^{1,2}, Daniela Colzani ^{1,2}, Elisa Borghi ^{1,2},
Gianvincenzo Zuccotti ^{7,8}, Mario C Raviglione ⁹, Elisabetta Tanzi ^{1,2},
Antonella Amendola ^{1,2}

To cite: Canuti M, Bianchi S, Kolbl O, *et al*. Waiting for the truth: is reluctance in accepting an early origin hypothesis for SARS-CoV-2 delaying our understanding of viral emergence? *BMJ Global Health* 2022;**7**:e008386. doi:10.1136/bmjgh-2021-008386

Handling editor Seye Abimbola

MC and SB contributed equally.

MC and SB are joint first authors.

Received 27 December 2021

Accepted 19 February 2022



© Author(s) (or their employer(s)) 2022. Re-use permitted under CC BY-NC. No commercial re-use. See rights and permissions. Published by BMJ.

For numbered affiliations see end of article.

Correspondence to

Dr Marta Canuti;
marta.canuti@gmail.com,
Dr Silvia Bianchi;
silvia.bianchi@unimi.it and
Professor Antonella Amendola;
antonella.amendola@unimi.it

ABSTRACT

Two years after the start of the COVID-19 pandemic, key questions about the emergence of its aetiological agent (SARS-CoV-2) remain a matter of considerable debate. Identifying when SARS-CoV-2 began spreading among people is one of those questions. Although the current canonically accepted timeline hypothesises viral emergence in Wuhan, China, in November or December 2019, a growing body of diverse studies provides evidence that the virus may have been spreading worldwide weeks, or even months, prior to that time. However, the hypothesis of earlier SARS-CoV-2 circulation is often dismissed with prejudicial scepticism and experimental studies pointing to early origins are frequently and speculatively attributed to false-positive tests. In this paper, we critically review current evidence that SARS-CoV-2 had been circulating prior to December of 2019, and emphasise how, despite some scientific limitations, this hypothesis should no longer be ignored and considered sufficient to warrant further larger-scale studies to determine its veracity.

INTRODUCTION

Tomorrow, when I wake, or think I do, what shall I say of today? –Waiting for Godot, Samuel Beckett.

During the spring of 2020, the sudden outburst of one of the most devastating pandemics of modern times changed our life, possibly forever. Almost 2 years after the beginning of this transition into the ‘new normal’, with enormous daily case and death counts, tight disease surveillance and lockdowns, there are still many unresolved questions about how this could have happened. Where did the virus come from? When, where and how did the virus jump the species barrier? When did the virus become highly capable of infecting humans? Did the worldwide

Summary box

- ▶ A growing body of studies provides evidence for the global circulation of SARS-CoV-2 prior to December 2019, contradicting the currently hypothesised timeline of the original viral emergence in Hubei province of China around November 2019; however, any suggestion of an earlier SARS-CoV-2 circulation is met with scepticism.
- ▶ Several studies performed independently by different groups retrospectively demonstrated the presence of antibodies and viral RNA in clinical samples and showed SARS-CoV-2 community circulation by detecting viral RNA in wastewater at times inconsistent with November 2019 emergence.
- ▶ Despite some limitations, combining the knowledge acquired from these studies is sufficient to warrant further larger-scale investigations to determine the veracity of this hypothesis.
- ▶ If proven true, an earlier than currently believed worldwide spread of SARS-CoV-2 will provide essential clues for understanding the genesis of this pandemic and offer invaluable lessons from our successes and failures with crucial implications for future pandemic preparedness and global health.

respiratory and general disease surveillance systems allow the virus to spread undetected? If history is an indication, some of these questions may remain unanswered for a long time to come.

Late in December 2019, Chinese clinicians began reporting that a pneumonia of unknown origin had affected people in Wuhan, Hubei, and on 31 December, the WHO became aware of the situation after identifying a ProMED post and online reports about the outbreak.¹ The WHO acted quickly, and on 5 January 2020, informed all

its Member States about the outbreak with the general reminder that ‘public health measures and surveillance of influenza and severe acute respiratory infections still apply’.² Quite rapidly, the genome of a novel coronavirus, later named SARS-CoV-2, was sequenced and determined to be the causative agent. Subsequently, health authorities realised that this new agent was capable of human-to-human transmission. As a result, the first border control measures for screening air passengers coming from Wuhan were implemented in many countries. On 30 January 2020, as recommended by its Emergency Committee, the WHO declared the outbreak ‘Public Health Emergency of International Concern’ and alerted countries to be prepared for the deployment of emergency measures such as containment, active surveillance and contact tracing.¹ However, by then, the virus had already been spreading worldwide³ and by 27 February 2020, cases of COVID-19 had already been reported in 53 countries, prompting the WHO on 11 March to characterise the disease as a global pandemic.¹

The origin of SARS-CoV-2 is still hotly debated and there is no dispositive proof of whether the virus started its spread after a single or multiple zoonotic events or if the virus ‘escaped’ from a research laboratory through accidental exposure or breach of safety protocols. So far, zoonotic emergence is considered the most likely option by some scientists.⁴ Some of the first COVID-19 cases in Wuhan were, in fact, linked to a seafood market that sold a wide array of live animals, and similar zoonotic origins have already been proven for other human coronaviruses.⁵ However, targeted scientific investigations concluded that the market might have simply acted as an amplification site.¹ Additionally, although close relatives of SARS-CoV-2 have been detected in animals, with bats harbouring the closest relatives to this virus, genomes sufficiently close to early SARS-CoV-2 isolates to act as zoonotic sources have not yet been identified in any species, making it difficult to pinpoint any potential intermediate hosts.^{1 4 6 7} Since genomically closest bat coronaviruses have been identified thousands of kilometres away from Wuhan and no SARS-CoV-2-infected animals have been linked to the wet markets in Wuhan, the question of how the progenitor virus arrived in Wuhan remains unanswered. Therefore, to clarify the mechanisms at the origin of this virus, identifying more accurately when and where the first human cases occurred is crucial to determine the original infection source(s).

EVIDENCE OF EARLIEST SARS-COV-2 INFECTIONS

Although considerable uncertainty around the early phases of the epidemic in Wuhan remains, the earliest patient with laboratory-confirmed COVID-19 reported in literature developed symptoms on 1 December 2019.⁸ However, several literature reports provide clues for virus circulation during that time and earlier in other parts of the world⁹ (table 1). In the USA, SARS-CoV-2 reactive antibodies were detected in over 100 blood samples collected

in several different states in early December 2019.¹⁰ In Brazil, environmental surveillance monitoring demonstrated early SARS-CoV-2 community spread at the end of November 2019 by detecting viral RNA in wastewater.¹¹ A study performed in the UK identified a few blood donors in May 2019 whose sera presented SARS-CoV-2 S-reactive antibodies associated with a presumed current immune response.¹² In France, antibodies against SARS-CoV-2 were found in serum samples collected in November 2019, and viral RNA was detected in December 2019 in a respiratory sample from a patient hospitalised for haemoptysis.^{13 14} However, most of the studies that investigated and found evidence for early SARS-CoV-2 circulation were performed in Italy, the first European country reporting sustained community transmission.

The first official and confirmed non-travel-related Italian COVID-19 case was identified on 20 February 2020, in Codogno, a small town located in Lombardy, the most populous region of Italy. Italy quickly became the epicentre of the COVID-19 European epidemic and by March 2020, it surpassed China in the number of officially reported cases. Within Italy, Lombardy was the most affected region during the first pandemic wave.^{1 15–18} During March–May 2020, Italy registered a 31.7% increase in all-cause deaths compared with the same period in the quinquennia 2015–2019 (excess deaths due to all causes), an increase to which the heaviest contributors were the regions from the north (+61.1%) and especially Lombardy, which experienced 111.8% increase.¹⁸ Lombardy was the first area of the Western World to be severely affected by the pandemic and one of the regions that suffered its heaviest consequences. These reasons are likely why many research groups were induced to retrospectively investigate the initial phases of the COVID-19 outbreak in Italy.

From the very beginning, it was clear that sustained viral spread started in Lombardy weeks before the first detection of the virus in the Codogno case. Indeed, the estimated net reproduction number had been above the epidemic threshold since late January 2020 and contact tracing demonstrated ongoing transmission throughout January 2020.^{19 20} A study performed later by the Italian National Institute of Health demonstrated that the virus was circulating in Milan (Lombardy) and Turin (Piedmont) by the end of 2019 in sufficient numbers to allow the detection of viral RNA in sewage water samples collected on 18 December 2019.²¹ Finally, several studies performed independently by different groups demonstrated the presence of antibodies against SARS-CoV-2 in blood samples since September 2019^{22–24} and of viral RNA in clinical samples collected from patients with cutaneous manifestations (dermatosis and measles-like rash) as early as 12 September 2019.^{24–26}

CRITICISMS AND SUPPORTING FACTS

The spread of SARS-CoV-2 outside of China prior to December 2019 directly contradicts the currently

Table 1 Considered retrospective studies that found evidence for an early SARS-CoV-2 circulation

Reference	Methods	Samples	Study period and location	Findings	Caaveats (#) and confirmations (§)
25	Nested-PCR, Sanger sequencing	39 oropharyngeal swabs collected within the measles/rubella surveillance system	Sep 2019–Feb 2020 Lombardy (Italy)	Viral RNA detected in Dec 2019	# The use of nested-PCR increases the chances of false positives # Single methodological approach § The positive sample was confirmed by sequencing § Samples were analysed in a follow-up study ²⁴
24	Nested-PCR, Sanger sequencing, commercial ELISA, neutralisation assay	Oropharyngeal swabs, urine and sera (N=435) collected within the measles/rubella surveillance system	Aug 2018–Apr 2021 Lombardy (Italy)	Viral RNA detected since 12 Sep 2019. The first positive patient was also IgG and IgM positive	# The use of nested-PCR increases the chances of false positives § Positive samples were confirmed by sequencing, and different variants were found § Samples from 100 patients collected before the first case were RNA negative § Follow-up studies with an external laboratory are currently ongoing
26	Immunohistochemistry, <i>in-situ</i> hybridisation	One skin biopsy from a patient with dermatosis	Nov 2019 Milan (Lombardy, Italy)	Presence of SARS-CoV-2 in paraffin samples	# Viral RNA undetected with RT-PCR technology # Results not confirmed by another laboratory § Positivity was confirmed by two different techniques on paraffin sections
21	Nested-PCR, RT-PCR	40 composite wastewater samples collected within the framework of wastewater-based environmental monitoring	Oct 2019–Feb 2020 Milan (Lombardy), Turin (Piedmont), Bologna (Emilia Romagna) (Italy)	15 samples were confirmed positive by both methods. The earliest detections were on 18 Dec 2019 (Milan and Turin) and 29 Jan 2020 (Bologna)	# The use of nested-PCR increases the chances of false positives # Results not confirmed by another laboratory § Positivity confirmed by RT-PCR and sequencing § The plants where the positive samples were collected serve a population of approximately 4 million inhabitants
20 49	Contact tracing from Directorate General for Health, Lombardy Region	Official records	Jan 2020–Feb 2020 Lombardy (Italy)	Epidemiological investigations suggested a sustained transmission across all Lombardy provinces from 1 Jan 2020	# Uncertainties due to difficulties in identifying the correct epidemiological links between cases # No data were available to distinguish imported <i>versus</i> local infections
22	In-house ELISA, virus neutralisation assay	959 blood samples from asymptomatic individuals enrolled in a prospective lung cancer screening trial	Sep 2019–Mar 2020 Italy	RBD-specific Ig in 11.6% samples since Sept 2019 (14%); cluster of positive cases (>30%) in the 2nd week of Feb 2020. Highest positive rate (53.2%) in Lombardy.	# No samples before September 2019 were analysed # Ig detection is less specific, and a proportion of cases could be attributed to false positivity § Six samples also showed neutralising activity § Results for a subsample confirmed in a follow-up study ²³
23	ELISA, microneutralisation assay	29 plasma samples	Jul 2019–Feb 2020 Italy	Presence of IgM and IgG antibodies in the pre-pandemic period	# Ig detection is less specific, and a proportion of cases could be attributed to false positivity § Results partially confirmed by an external laboratory
13	Commercial ELISA, in-house virus microneutralisation assay	Serum samples collected from 9144 adults from a French general population-based cohort	Nov 2019–Mar 2020 France	3.9% of samples were positive to anti-SARS-CoV-2 IgG test and 13 had neutralising activities	# No samples before November 2019 were analysed # Ig detection is less specific, and a proportion of cases could be attributed to false positivity # Results not confirmed by another laboratory § Thirteen samples showed neutralising activity
14	RT-PCR	14 respiratory samples collected from patients hospitalised in the intensive care unit	Dec 2019–Jan 2020 France	Confirmed diagnosis of SARS-CoV-2 infection in one patient with haemoptysis in Dec 2019	# Positivity was not confirmed by sequencing # Single methodological approach # Results not confirmed by another laboratory § The result was confirmed with a second RT-PCR protocol and by different laboratory operators

Continued

Table 1 Continued

Reference	Methods	Samples	Study period and location	Findings	Caveats (#) and confirmations (§)
12	Flow cytometry-based method	101 blood samples from uninfected individuals	May 2019 UK	Presence of pre-existing antibodies recognising SARS-CoV-2 in uninfected individuals	# Ig detection is less specific, and a proportion of cases could be attributed to false positivity # Results not confirmed by another laboratory § Two donors presented simultaneously antibodies of the three classes
11	RT-PCR, SISPA, metagenomics	Six human sewage water samples	Oct 2019–Mar 2020 Santa Catarina (Brazil)	Viral RNA detected since 27 Nov 2019	# No clinical records of local COVID-19 cases in 2019 # The sewer where the samples were collected serves a population of approximately 5000 inhabitants § Positive samples were confirmed by sequencing § Virus detection was confirmed by an independent laboratory
10	Pan-Ig ELISA, microneutralisation, ortho total Ig S1 ELISA, RBD/ACE2 blocking activity assays	7389 serum specimens from blood donations	Dec 2019–Jan 2020 California, Connecticut, Iowa, Massachusetts, Michigan, Oregon, Rhode Island, Washington, Wisconsin (USA)	1.4% of samples were reactive by pan-Ig, 84 of 90 had neutralising activity. Presence of anti-SARS-CoV-2 reactive antibodies since 13–16 Dec 2019.	# No samples before Dec 2019 were analysed # Ig detection is less specific, and a proportion of cases could be attributed to false positivity # Results not confirmed by another laboratory § The majority of Ig-positive samples showed neutralising activity

Ig, immunoglobulin; RBD, receptor-binding domain of SARS-CoV-2; RT-PCR, real-time PCR; S1, subunit 1 of the spike protein of SARS-CoV-2; SISPA, sequence-independent single-primer amplification.

hypothesised timeline of the original viral emergence in China’s Hubei province around November 2019^{3 27 28} and its subsequent introduction in Europe and North America in January 2020.^{29 30} However, sequence-based relaxed molecular clock methods used to date ancestral nodes in phylogenetic trees are influenced by sequence availability and temporal signal present in the data. The addition of even a few early strains, which may have gone undetected at the beginning, can significantly shift the relative likelihoods of these predictions.^{24 31} More recent computational analyses including a very large number of complete genomes have moved the time of viral emergence to well before the major Wuhan outbreak, up to the summer of 2019.^{32–34}

Because of these disagreements, laboratory evidence for early circulation is often dismissed and labelled as a result of false-positive testing. Antibody detection results can indeed be affected by the presence in sera of antibodies which, although able to recognise SARS-CoV-2 antigens, were induced by other agents.¹² However, the presence of SARS-CoV-2 neutralising activity in these sera and the fact that several patients presented more than one class of antibodies recognising SARS-CoV-2 suggest that, although some cross-reactivity should be taken into account, at least some of the sera could contain antibodies induced by a prior SARS-CoV-2 infection.^{10 12 22–24}

Direct evidence for early viral circulation can be obtained by detecting SARS-CoV-2 RNA. Techniques used for this purpose, especially if accompanied by sequencing, are more specific and less influenced by previous infections due to other pathogens. However, PCR-based methods are highly sensitive and, therefore,

more prone to false-positive results. Various methods have been used to detect SARS-CoV-2 RNA in 2019 samples collected outside of China, including RNA fluorescence *in-situ* hybridisation,²⁶ RT-qPCR^{11 14 21} and nested-PCR,^{21 24 25} and most of the results obtained were confirmed by sequencing.^{11 21 24 25} Unfortunately, none of these studies could recover complete genomic sequences needed for phylogenetic and coalescence analyses. This is likely a problem tracing back to the unsuitability of the available samples, which did not contain high enough concentrations of high-quality RNA. Even in clinical samples found to be RNA positive, viral load was low, implying that only a few viral templates would be available for amplification. For instance, the positive oropharyngeal and urine samples were collected within the framework of measles surveillance from patients who developed a skin rash,^{24 25} which may appear late during infection, when respiratory symptoms have been resolved, or even before the onset of other COVID-19 symptoms,^{35 36} resulting in low viral loads in swabs collected at the moment of the rash onset. Interestingly, one study detected multiple positive patients, some of whom also presented SARS-CoV-2-recognising antibodies, and observed polymorphisms in viral sequences, a factor that excludes laboratory contamination by a single positive control sequence.²⁴

IMPLICATIONS OF AVAILABLE EARLY CIRCULATION EVIDENCE

Some studies that retrospectively screened respiratory samples collected from patients with influenza-like and respiratory symptoms, including those collected within

Table 2 Considered retrospective studies that found no evidence for an early SARS-CoV-2 circulation

Reference	Methods	Samples	Study period and location	Findings	Caveats (#) and confirmations (§)
³⁸	RT-PCR	631 nasopharyngeal swabs collected from outpatients with ILI by sentinel physicians within the regional influenza surveillance network	Nov 2019–Feb 2020 Lombardy (Italy)	No evidence of SARS-CoV-2 circulation	# Single methodological approach # Results not confirmed by another laboratory
³⁷	RT-PCR	1581 respiratory samples collected within the framework of the regional influenza surveillance system	Nov 2019–Apr 2020 Lombardy (Italy)	First SARS-CoV-2 detection at the beginning of Mar	# Single methodological approach # Results not confirmed by another laboratory
³⁹	RT-PCR	Mucus obtained from nasopharyngeal swabs collected from 166 patients with SARI	Nov 2019–Mar 2020 Rome (Italy)	No evidence of SARS-CoV-2 circulation	# Patients from an area not primarily involved during the first wave # Single methodological approach # Results not confirmed by another laboratory
⁴¹	RT-PCR	1683 oropharyngeal swabs from hospitalised patients with SARI	Nov 2019–Feb 2020 Liguria, Tuscany, Lazio, Puglia (Italy)	No evidence of SARS-CoV-2 circulation	# Patients from an area not primarily involved during the first wave # Single methodological approach # Results not confirmed by another laboratory
⁴⁰	RT-PCR, metagenomic sequencing	5833 nasopharyngeal and oropharyngeal swabs collected from patients hospitalised for ILI	Nov 2019–Mar 2020 Valencia (Spain)	No evidence of SARS-CoV-2 circulation before Mar 2020	# Patients from an area not primarily involved during the first wave # Single methodological approach # Negative results not confirmed by another laboratory § The positive sample from March was confirmed by sequencing § The positivity was confirmed by another laboratory

ILI, influenza-like illness; RT-PCR, real-time PCR; SARI, severe acute respiratory illness.

the framework of influenza surveillance, found no evidence for SARS-CoV-2 (table 2),^{37–41} including during times when community SARS-CoV-2 transmission in respective populations was already occurring.²⁰ Nonetheless, the combination of all knowledge acquired so far from retrospective studies strongly suggests that SARS-CoV-2 was circulating outside of China considerably earlier than the currently postulated time frame of late December 2019/early January 2020, at least in some parts of the world.

Good epidemiological insight comes from sewage water testing. For the virus to be detectable in wastewater, a somewhat sustained virus circulation must be happening, as demonstrated in a recent publication performing environmental surveillance in Lombardy.⁴² Fongaro *et al* showed that SARS-CoV-2 load remained temporally constant in the wastewater of Santa Catarina, Brazil, but increased in March 2020 in concomitance with the surge in COVID-19 cases.¹¹ Results from La Rosa *et al*, who detected the virus in the wastewater of Milan and Turin in December 2019,²¹ are supported by the detection of SARS-CoV-2 in clinical samples since September 2019²⁴ and by the subsequent steady increase in COVID-19 cases observed at the beginning of 2020,²⁰ before the exponential growth in cases became evident. However, it remains puzzling why the virus was detected in samples collected within the framework of measles surveillance and not in those available to the influenza surveillance network.

This incongruence could be explained by the fact that skin manifestations draw more attention than respiratory signs and are almost always reported, making measles surveillance more sensitive and comprehensive, while also monitoring a lower and more manageable number of cases. In fact, mild COVID-19 cases may have been masked by the ongoing influenza season⁴³ and the high number of respiratory infections common to this time of the year.

Although the notion of a somewhat sustained but unnoticed viral circulation may be difficult to accept after so many deaths and so many months of pandemic viral circulation, this possibility is ‘not astonishing’, as recently noted by Petti.⁹ Certainly, the virus had been circulating for some time before such dramatic excess mortality could have become noticeable and a low number of COVID-19-related deaths could have gone unnoticed among the high number of deaths associated with pneumonia of unknown origin that occur every year. Although currently difficult to establish without extensive sequencing of early strains, a virus with reduced transmissibility and/or virulence could explain a slow, undetected diffusion during the early months of its spread, causing only sporadic cases and/or limited outbreaks.⁴⁴ The emergence of more infectious strains occurred several times during this pandemic,⁴⁵ and differences in replication dynamics have been demonstrated *in vitro* for some early strains.⁴⁶ However, preliminary sequencing

results suggest that the strain circulating in pre-pandemic Europe in late 2019 may have been already capable of efficient human-to-human transmission²⁴ and it has been recently hypothesised that SARS-CoV-2 strains may have acquired adaptive mutations in Europe, while spreading in parallel with Asian strains.⁴⁷ Scattered circulation combined with the lack of awareness may have contributed to a slow and undetected early spread.^{44 48 49} These are aspects that future studies should investigate by retrospectively identifying, sequencing and studying *in vitro* early circulating strains.

To determine whether SARS-CoV-2 was already spreading outside of China in 2019, it is crucial to broaden our efforts and consider a wider geographical area and a larger timespan when investigating viral emergence. In particular, future research should focus on samples where a higher viral load can be expected and thus that are more suitable for virus detection and sequencing, that is, on severe cases of patients with pneumonia admitted to intensive care units.

SCEPTICISM AND RELUCTANCE TO CONSIDER THE EARLY ORIGIN HYPOTHESIS

An analysis of the information transmission chain regarding the hypothesis of early circulation outside China reveals not only a dismissive attitude towards this hypothesis,⁵⁰ but also that this hypothesis is linked to other hypotheses regarding SARS-CoV-2. In other words, accepting the idea that SARS-CoV-2 or its progenitor(s) might have circulated in many regions of the world for months before it was discovered in Wuhan challenges several widely accepted assumptions about this virus. However, it is important to note that the early origin of SARS-CoV-2 has no bearing on the debate about the laboratory leak *versus* natural origin, and it does not exclude the possibility of its origin in Hubei or somewhere else in China. The acceptance of the early origin is inconvenient in that we may no longer be able to use the circumstantial evidence of time and location of the first detection of SARS-CoV-2 as our final answer.

Despite the increasing documentation available in support of its early circulation, current scientific literature discussing the origin of SARS-CoV-2 is almost exclusively focused on the November/December 2019 hypothesis, completely ignoring this growing body of contradictory evidence. In fact, the possibility of early circulation is only seldom mentioned or discussed in such papers. Furthermore, as this alternative hypothesis clearly contradicts the timeline that is today held as the most likely, when these studies are cited, it is done dismissively, minimising the results obtained by numerous independent research groups. This attitude, pervasive among high-ranking journals, clearly demonstrates scepticism and has the consequence of avoiding a more critical interpretation of scientific data and of discouraging a constructive scientific debate that should consider all available facts when advancing a hypothesis and re-evaluate assumptions in

light of new evidence. Additionally, this bias often results in rejection of manuscripts in support of an early SARS-CoV-2 circulation, reinforcing the ‘echo chamber’ effect. Science is a quest for ultimate truth, which shall not be discouraged by such mindset.

CONCLUSIONS

Research into the origins of SARS-CoV-2 is a challenging and fraught undertaking and there is still much that needs to be elucidated. Each study providing evidence for early circulation of SARS-CoV-2 might look inconclusive, but combining all data together reveals an emerging pattern. As all it would take to establish early circulation is a single confirmed positive case, attributing all results to false positives quickly becomes probabilistically untenable. Nonetheless, given the critical implications of these findings, it is important to obtain confirmatory proof for an early viral spread or lack thereof by independent investigations performed in WHO-accredited laboratories, as already occurred in one instance.²³ Rescreening with more standardised and sensitive methods will provide confirmation for studies that found evidence for an early circulation and for those that did not. This aspect is crucial especially in those geographical areas affected during the first wave and where the lack of early viral detection is in obvious conflict with overwhelming epidemiological evidence. Additional approaches—such as metagenomic sequencing—could also be used to obtain more sequence information, which is essential for dating the beginning of viral spread more accurately. Finally, a more systematic approach to retrospectively test for anti-SARS-CoV-2 antibodies in numerous serum samples collected from a broad area could allow the identification of seroprevalence peaks that would help filter out potential background noise caused by cross-reactivity. International public health authorities should ideally coordinate such studies.

SARS-CoV-2 has cost the world the lives of millions of people, and the next (inevitable) pandemic might have more devastating outcomes. To be better prepared in the future and timely identify the emergence of novel pathogens, it is crucial to fully understand this pandemic and learn from our successes and our failures. Despite the technical limitations of available early origin studies, even a remote possibility that positive tests indicate an early SARS-CoV-2 circulation should be considered sufficient to warrant the scaling up of research to more samples from more regions and through a wider timespan. Time is running out: valuable samples that may contain the key to the understanding of SARS-CoV-2 origin might already have been destroyed as their regulatory storage time requirements lapse. Many more will meet the same fate in the coming months and years. What is there to lose in accepting this hypothesis as tenable and exploring it urgently before the chances of finding the answers to explain how this pandemic emerged are gone forever?

...Let us not waste our time in idle discourse! Let us do something, while we have the chance... at this place, at this moment of time, all mankind is us, whether we like it or not. Let us make the most of it before it is too late! –Waiting for Godot, Samuel Beckett.

Author affiliations

- ¹Department of Health Sciences, Università degli Studi di Milano, Milan, Italy
²Coordinated Research Center “EpiSoMI”, Università degli Studi di Milano, Milan, Italy
³Faculty of Arts, University of Lausanne, Lausanne, Switzerland
⁴Department of Biology, Temple University, Philadelphia, Pennsylvania, USA
⁵Institute for Genomics and Evolutionary Medicine, Temple University, Philadelphia, Pennsylvania, USA
⁶Center for Excellence in Genome Medicine and Research, King Abdulaziz University, Jeddah, Saudi Arabia
⁷Department of Pediatrics, Ospedale dei Bambini, Università degli Studi di Milano, Milan, Italy
⁸Romeo and Enrica Invernizzi Pediatric Research Center, Università degli Studi di Milano, Milan, Italy
⁹Centre for Multidisciplinary Research in Health Science, Università degli Studi di Milano, Milan, Italy

Twitter Otto Kolbl @OttoKolbl

Contributors MC, SB and AA conceived the initial idea for the manuscript, and MC wrote the first draft. All authors provided critical feedback and approved the final submission.

Funding This study was funded by Romeo and Enrica Invernizzi Pediatric Research Center, Università degli Studi di Milano, Milan, Italy.

Competing interests None declared.

Patient consent for publication Not required.

Ethics approval This study does not involve human participants.

Provenance and peer review Not commissioned; externally peer reviewed.

Data availability statement There are no data in this work.

Open access This is an open access article distributed in accordance with the Creative Commons Attribution Non Commercial (CC BY-NC 4.0) license, which permits others to distribute, remix, adapt, build upon this work non-commercially, and license their derivative works on different terms, provided the original work is properly cited, appropriate credit is given, any changes made indicated, and the use is non-commercial. See: <http://creativecommons.org/licenses/by-nc/4.0/>.

ORCID iDs

- Marta Canuti <https://orcid.org/0000-0002-9959-128X>
 Silvia Bianchi <https://orcid.org/0000-0002-1365-9408>
 Sergei L Kosakovskiy <https://orcid.org/0000-0003-4817-4029>
 Sudhir Kumar <https://orcid.org/0000-0002-9918-8212>
 Maria Gori <https://orcid.org/0000-0001-6478-2791>
 Clara Fappani <https://orcid.org/0000-0002-1573-5531>
 Daniela Colzani <https://orcid.org/0000-0002-4079-1957>
 Elisa Borghi <https://orcid.org/0000-0002-1893-0455>
 Gianvincenzo Zuccotti <https://orcid.org/0000-0002-2795-9874>
 Mario C Raviglione <https://orcid.org/0000-0002-9331-2067>
 Elisabetta Tanzi <https://orcid.org/0000-0002-4119-701X>
 Antonella Amendola <https://orcid.org/0000-0002-0499-8977>

REFERENCES

- Singh S, McNab C, Olson RM, *et al*. How an outbreak became a pandemic: a chronological analysis of crucial junctures and international obligations in the early months of the COVID-19 pandemic. *Lancet* 2021;398:2109–24.
- Pneumonia of unknown cause – China, 2020. Available: <https://www.who.int/emergencies/disease-outbreak-news/item/2020-DON229>
- Kumar S, Tao Q, Weaver S. An evolutionary portrait of the progenitor SARS-CoV-2 and its dominant offshoots in COVID-19 pandemic. *Mol Biol Evol* 2021.
- Holmes EC, Goldstein SA, Rasmussen AL, *et al*. The origins of SARS-CoV-2: a critical review. *Cell* 2021;184:4848–56.
- Cui J, Li F, Shi Z-L. Origin and evolution of pathogenic coronaviruses. *Nat Rev Microbiol* 2019;17:181–92.
- Leitner T, Kumar S. Where did SARS-CoV-2 come from? *Mol Biol Evol* 2020;37:2463–4.
- Temmam S, Vongphayloth K, Baquero Salazar E. Coronaviruses with a SARS-CoV-2-like receptor-binding domain allowing ACE2-mediated entry into human cells isolated from bats of Indochinese Peninsula. *Research Square*.
- Huang C, Wang Y, Li X, *et al*. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *The Lancet* 2020;395:497–506.
- Petti S. Undetected and relatively sustained severe acute respiratory syndrome coronavirus 2 circulation worldwide during 2019. *Clin Infect Dis* 2021;395.
- Basavaraju SV, Patton ME, Grimm K, *et al*. Serologic Testing of US Blood Donations to Identify Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2)-Reactive Antibodies: December 2019–January 2020. *Clin Infect Dis* 2021;72:e1004–9.
- Fongaro G, Stoco PH, Souza DSM, *et al*. The presence of SARS-CoV-2 RNA in human sewage in Santa Catarina, Brazil, November 2019. *Sci Total Environ* 2021;778:146198.
- Ng KW, Faulkner N, Cornish GH, *et al*. Preexisting and de novo humoral immunity to SARS-CoV-2 in humans. *Science* 2020;370:1339–43.
- Carrat F, Figoni J, Henny J, *et al*. Evidence of early circulation of SARS-CoV-2 in France: findings from the population-based “CONSTANCES” cohort. *Eur J Epidemiol* 2021;36:219–22.
- Deslandes A, Berti V, Tandjaoui-Lambotte Y, *et al*. SARS-CoV-2 was already spreading in France in late December 2019. *Int J Antimicrob Agents* 2020;55:106006.
- Grasselli G, Pesenti A, Cecconi M. Critical care utilization for the COVID-19 outbreak in Lombardy, Italy: early experience and forecast during an emergency response. *JAMA* 2020;323:1545.
- Tosi D, Chiappa M. Understanding the geographical spread of COVID-19 in relation with goods regional routes and governmental Decrees: the Lombardy region case study. *SN Comput Sci* 2021;2:203.
- Odone A, Delmonte D, Scognamiglio T, *et al*. COVID-19 deaths in Lombardy, Italy: data in context. *Lancet Public Health* 2020;5:e310.
- ISTAT, ISS. Impatto dell’epidemia COVID-19 sulla mortalità totale della popolazione residente anno 2020, 2021. Available: https://www.istat.it/it/files/2021/03/Report_ISS_Istat_2020_5_marzo.pdf
- Riccardo F, Ajelli M, Andrianou XD, *et al*. Epidemiological characteristics of COVID-19 cases and estimates of the reproductive numbers 1 month into the epidemic, Italy, 28 January to 31 March 2020. *Euro Surveill* 2020;25:2000790.
- Cereda D, Tirani M, Rovida F. The early phase of the COVID-19 outbreak in Lombardy, Italy. *arXiv* 2020.
- La Rosa G, Mancini P, Bonanno Ferraro G, *et al*. SARS-CoV-2 has been circulating in northern Italy since December 2019: evidence from environmental monitoring. *Sci Total Environ* 2021;750:141711.
- Apolone G, Montomoli E, Manenti A, *et al*. Unexpected detection of SARS-CoV-2 antibodies in the pre-pandemic period in Italy. *Tumori* 2021;107:0300891620974755
- Montomoli E, Apolone G, Manenti A, *et al*. Timeline of SARS-CoV-2 spread in Italy: results from an independent serological retesting. *Viruses* 2021;14:61.
- Amendola A, Canuti M, Bianchi S, *et al*. Molecular evidence for SARS-CoV-2 in samples collected from patients with Morbilliform eruptions since late summer 2019 in Lombardy, Northern Italy. *SSRN Electronic Journal* 2021.
- Amendola A, Bianchi S, Gori M, *et al*. Evidence of SARS-CoV-2 RNA in an oropharyngeal swab specimen, Milan, Italy, early December 2019. *Emerg Infect Dis* 2021;27:648–50.
- Gianotti R, Barberis M, Fellegara G, *et al*. COVID-19-related dermatosis in November 2019: could this case be Italy’s patient zero? *Br J Dermatol* 2021;184:970–1.
- Pekar J, Worobey M, Moshiri N, *et al*. Timing the SARS-CoV-2 index case in Hubei Province. *Science* 2021;372:412–7.
- Duchene S, Featherstone L, Haritopoulou-Sinaniadou M, *et al*. Temporal signal and the phylodynamic threshold of SARS-CoV-2. *Virus Evol* 2020;6:veaa061.
- Worobey M, Pekar J, Larsen BB, *et al*. The emergence of SARS-CoV-2 in Europe and North America. *Science* 2020;370:564–70.
- Zehender G, Lai A, Bergna A, *et al*. Genomic characterization and phylogenetic analysis of SARS-CoV-2 in Italy. *J Med Virol* 2020;92:1637–40.
- Bloom JD. Recovery of deleted deep sequencing data sheds more light on the early Wuhan SARS-CoV-2 epidemic. *Mol Biol Evol* 2021;38:5211–24.

- 32 Caraballo-Ortiz MA, Miura S, Sanderford M, *et al.* TopHap: rapid inference of key phylogenetic structures from common haplotypes in large genome collections with limited diversity. *bioRxiv* 2021. doi:10.1101/2021.12.13.472454. [Epub ahead of print: 14 Dec 2021].
- 33 Xia X. Dating the common ancestor from an NCBI tree of 83688 high-quality and full-length SARS-CoV-2 genomes. *Viruses* 2021;13:1790.
- 34 Chiara M, Horner DS, Gissi C, *et al.* Comparative genomics reveals early emergence and biased spatiotemporal distribution of SARS-CoV-2. *Mol Biol Evol* 2021;38:2547–65.
- 35 Daneshgaran G, Dubin DP, Gould DJ. Cutaneous manifestations of COVID-19: an evidence-based review. *Am J Clin Dermatol* 2020;1–13.
- 36 Genovese G, Moltrasio C, Berti E, *et al.* Skin manifestations associated with COVID-19: current knowledge and future perspectives. *Dermatology* 2021;237:1–12.
- 37 Giardina F, Galli C, Pellegrinelli L, *et al.* No evidence of SARS-CoV-2 circulation in the framework of influenza surveillance between October 2019 and February 2020 in Lombardy, Italy. *Travel Med Infect Dis* 2021;40:102002.
- 38 Galli C, Pellegrinelli L, Bubba L, *et al.* When the COVID-19 pandemic surges during influenza season: lessons learnt from the sentinel laboratory-based surveillance of influenza-like illness in Lombardy during the 2019–2020 season. *Viruses* 2021;13:695.
- 39 Capalbo C, Bertamino E, Zerbetto A, *et al.* No evidence of SARS-CoV-2 circulation in Rome (Italy) during the Pre-Pandemic period: results of a retrospective surveillance. *Int J Environ Res Public Health* 2020;17:8461.
- 40 Mira-Iglesias A, Mengual-Chuliá B, Cano L. Retrospective screening for SARS-CoV-2 among influenza-like illness hospitalizations: 2018–2019 and 2019–2020 seasons, Valencia region, Spain. *Influenza other resp* 2021.
- 41 Panatto D, Orsi A, Pennati BM, *et al.* No evidence of SARS-CoV-2 in hospitalized patients with severe acute respiratory syndrome in five Italian hospitals from 1st November 2019 to 29th February 2020. *PLoS One* 2021;16:e0260947.
- 42 Castiglioni S, Schiarea S, Pellegrinelli L, *et al.* SARS-CoV-2 RNA in urban wastewater samples to monitor the COVID-19 pandemic in Lombardy, Italy (March–June 2020). *Sci Total Environ* 2022;806:150816.
- 43 Percivalle E, Cambiè G, Cassaniti I, *et al.* Prevalence of SARS-CoV-2 specific neutralising antibodies in blood donors from the Lodi red zone in Lombardy, Italy, as at 06 April 2020. *Euro Surveill* 2020;25:2001031.
- 44 Ruan Y, Wen H, He X, *et al.* A theoretical exploration of the origin and early evolution of a pandemic. *Sci Bull* 2021;66:1022–9.
- 45 Tian D, Sun Y, Zhou J, *et al.* The global epidemic of SARS-CoV-2 variants and their mutational immune escape. *J Med Virol* 2021. doi:10.1002/jmv.27376. [Epub ahead of print: 05 Oct 2021].
- 46 Yao H, Lu X, Chen Q, *et al.* Patient-derived SARS-CoV-2 mutations impact viral replication dynamics and infectivity in vitro and with clinical implications in vivo. *Cell Discov* 2020;6:1–16.
- 47 Ruan Y, Wen H, Hou M, *et al.* The twin-beginnings of COVID-19 in Asia and Europe – one prevails quickly. *National Science Review* 2021;nwab223.
- 48 Ylli A, Wu YY, Burazeri G, *et al.* The lower COVID-19 related mortality and incidence rates in eastern European countries are associated with delayed start of community circulation. *PLoS One* 2020;15:e0243411.
- 49 Cereda D, Manica M, Tirani M, *et al.* The early phase of the COVID-19 epidemic in Lombardy, Italy. *Epidemics* 2021;37:100528.
- 50 Kolbl O. *Analysis of the information transmission chain in academic research reveals reluctance to accept early circulation of SARS-CoV-2*, 2021.