

Investigating the Undetected Early European Covid-19 Outbreak between December 2019 and February 2020 and the Failures of Public Policy to Manage the Epidemic

May 2020

By Prof Dr Peer Ederer

Global Food and Agribusiness Network

The author can be contacted on LinkedIn

All documentation can also be accessed on the author's homepage

www.foodandagribusiness.org

Supplement III

Phylogenetic Mutation History Analysis

Supplement III

Phylogenetic Mutation History Analysis

Subjects in Supplement III:

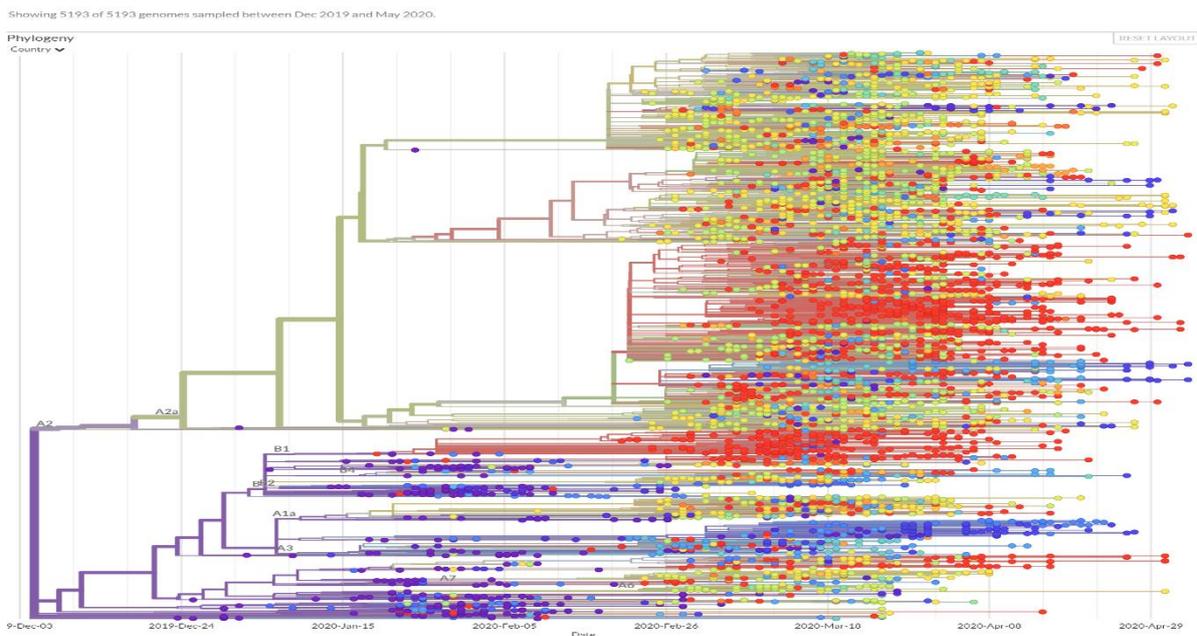
- a) Explanations on the phylogenetic dendrogram
- b) More details on the proposal for how the epidemic spread in its early days in Europe

III a) Explanations on the phylogenetic dendrogram

All charts mentioned in this text refer to the interactive dashboard which can be investigated on nextstrain.org, where the dendrogram looked like Figure S III a) on 14 May. The dendrograms on this dashboard can be configured in many different ways and presentations. The particular dendrogram used here, grows from left to right by day, and every clade is colored by the country in which this mutation is found. Purple stands for China, blue for Asia, green/yellow for Europe and red for Americas. The dates on the x-axis are calibrated on the day of discovery of the mutations. The corresponding infection date would usually be between 10 and 14 days earlier than that. All charts in the following text, are shown in Supplement IV.

Calculating backwards, on the basis of the relatively well established constant rate of mutations of the virus, the researchers who are building and analyzing the nextstrain.org dendrogram believe that the first fork of mutations with surviving descendants could have been registered during the first week of December 2019 in China, meaning the infection would have occurred at around the third week of November. The actually first known registration was sequenced on 24 December in Wuhan. This means that the first few weeks of the genetic evolution can only be reconstructed, but that this was not directly observed.

Figure S III a) SARS-CoV-2 phylogenetic dendrogram (= Chart 1 in supplement IV)

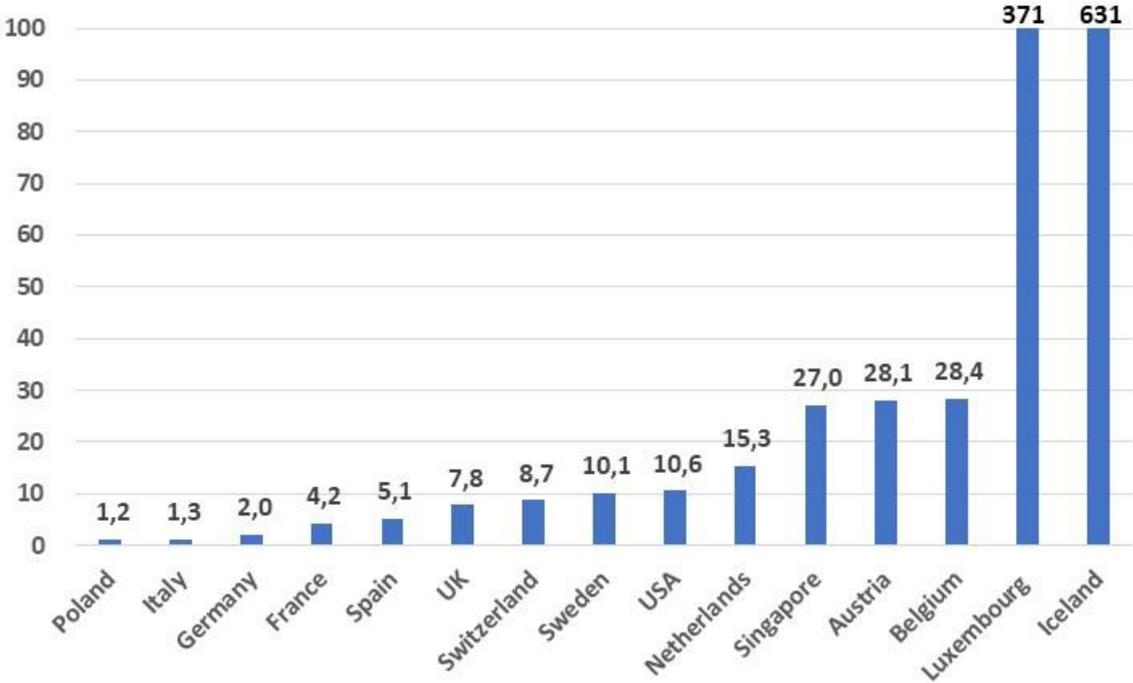


Source: [Nextstrain.org](https://nextstrain.org)

The dendrogram depends upon genetic sequencing analyses being conducted. According to the data shown on nextstrain.org as per 14 May, Iceland and Luxembourg had done the most of these on a per capita basis, at 631 and 371 per million inhabitants, respectively. Among larger countries, Belgium, Austria and Singapore are the busiest with around 28 each, followed by USA and Sweden at around 10 each, then France and Spain at 5 each, finally Germany, Italy and Poland with less than 2 each (Figure S III b). It is unfortunate that those countries who are at the center of interest for how the virus established itself in Europe, have not been committed to sequencing the mutations and contributing them to the database. The Democratic Republic of Congo has contributed more mutation sequences than Italy, and is not far behind Germany. However, there is nonetheless significant indirect evidence that can be gained from the phylogenetic dendrogram, which shows that the early European outbreak should have taken place in Bavaria in Germany.

The GISAID Team which procures the nextstrain.org site gave this important warning on 13 April: “It is important to note that there is currently not enough data from the early outbreak period to interpret the early history of global transmissions from few genomes in detail. Links that seem directly connected now are likely to be connected via other cases also from other countries not sampled and sometimes can be connected differently later with more data.”¹ Since that statement the number of submitted sequences has tripled to almost 40,000 at the end of May, but no new insights could be generated about the early transmission chains until March. That picture appears to be stabilizing.

Figure S III b) Sequencing of mutations per million inhabitants for selected countries



Source: Own calculations based on data of Nextstrain.org

III a 1) How the Chinese outbreak is displayed on the dendrogram

Chart 2) of Supplement IV zooms into the main Chinese clade and shows how according to the phylogenetic dendrogram, and as would be typical for a virus, it quickly generated many clades (branches). Every line on the chart represents one mutation being carried by one person. The longer a line is, the more genetic distance has been created without a fork that came to be known, which means that this line represents subsequent infections in human-to-human transmission. Clades remain unknown because either no member of them were sequenced (yet), or because they faded out on their own and are not prevalent anymore. Every fork usually corresponds to a time when this strain split into two or more separate paths with different mutations, even though sometimes the display algorithm also shows a fork with the same mutation if its members were found at different times and places. Every dot corresponds to the virus of one patient being registered and sequenced, and all known details about this person are displayed (age, location, exposure, genetic distance of the virus, its mutations, researcher, laboratory, etc).

Chart 2) shows how by the end of December 2019 (adjusting for the date of infection rather than the date of registration) there were already hundreds of clades running their course. Since all of the earliest registered mutations came from Wuhan, it is generally assumed that this is where the first fork of a human-to-human transmission occurred, but there is no direct evidence of that. The first original infections almost certainly have been much earlier than that, and indeed not even in Wuhan, nor even necessarily in China. But they would have left no descendants and had faded out on their own so that they could not be registered.

The one single mutation which became the ancestor for the current 2020 outbreak is the one that forked in some place at the end of November and led to the first Wuhan infections being registered in the third week of December. This makes Wuhan or China a possible place of this first patient 0 transmission, but by no means the certain place.

It is also usually assumed that the first patient 0 was a trader at the Wuhan Huanan Seafood Wholesale Market and that the virus jumped from an animal to humans there. The basis for this theory is that out of 585 swabs taken at locations in the market, 33 tested positive for SARS-CoV-2, all of which were found only in the area of the wild animal section of the market ². Also, 26 out of 47 confirmed patient cases before 01 January had an exposure history to the Huanan Seafood Wholesale Market. But critically, out of the first six cases known until 17 December, only two had an exposure to this market ³. It is more probable that the market just simply was the first super-spreading event for the virus ⁴. One of the earliest known infected patients was a trader at the market who sold shrimp ⁵. It might have been that her customers became infected, but this does not mean that she acquired the virus on the market. The smears at the wild animal section may have been created by an infected stray dog or cat.

The equally likely, if not even more likely possibility is that the first event at the end of November was a business meeting, a conference, a seminar, trade show or similar, where there were several people in the room, one of whom was infected and thus spread the disease. The reason that this is more likely, is because possibly either non-Chinese were present at that meeting, or Chinese who have frequent international contact. Without more primary data this cannot be known. This still does not answer where the virus came from, but it is clearly too simplistic to assume that the wild animal market of Wuhan was its origin, just because it was found there first. No animal that comes even close to being a vector could yet be identified. The closest bat

corona virus that is known to SARS-CoV-2, has a genetic distance of between 20 to 50 years⁶.

Chart 2) also shows, how by the end of December, the majority of known clades had already established themselves outside of Wuhan in other Chinese provinces. By the time the authorities locked down Wuhan and Hubei province on 23 January, the virus had already spread itself into thousands of places across all of China.

III a 2) Two super-spreading events explained by the phylogenetic dendrogram

Chart 3) zooms further into the Chinese main clade and demonstrates how a super-spreader event might look like in the phylogenetic dendrogram. If there is a long vertical line, then one person had been infecting several dozen or hundreds of persons on the same day, and thus established dozens of separate infection chains. The long horizontal clade lines do not represent a person who was directly infected at that event, but an infection chain that originated at that event.

The dendrogram imputes the event shown on chart 3) for 04 January, so it should have taken place 10 to 14 days earlier, between 21 and 25 December. Many foreigners must have visited this event, because there is an outbreak in Spanish Valencia discovered at the end of February which links directly back to this event, as well as other outbreaks in Thailand, Saudi Arabia and Australia. Many more countries show clades which forked shortly afterwards, and also many different cities in China were directly infected there.

One candidate for such an event could be the grand Light Festival which opened in Wuhan on 24 December, and which was a China-wide and international tourist attraction. Maybe there was one person at the Light Festival who had frequent contact with visitors and who was infectious with covid-19. This could have been for instance a street stall food vendor. Other candidates could be one of the many Christmas parades that were conducted across the city. Whatever this event was, it established what became known as the B clade of the virus. There are 29 subsequent branches which ultimately became known all over the world to have been direct descendants of this one super-spreading event, according to the dendrogram. There were probably several hundred or maybe even thousands of additional infection chains in China or in Wuhan itself which originated at that event, but did not become registered, or in the meanwhile faded out.

Chart 4) zooms further into the first so-called patient 0 in the United States. The story of this 35-year old man is well documented⁷. He visited his family in Wuhan and returned to the Snohomish county north of Seattle in US Washington State on 15 January on a 20-hour flight connection⁸. He then travelled on a shared public transport to his home town. He had a slight cough and later on fever. Because he had seen a health alert from the American CDC, he decided to see a health care provider. On 20 January he was confirmed positive with covid-19. The patient was immediately isolated and extensive contact tracing work was initiated to find everybody who he might have met before presenting to the clinic. Despite dozens of tests and requested isolations, not a single other person was known to have been infected at that time. The authorities believed that an outbreak had been prevented.

The next positive case in Washington State was confirmed five weeks later on 28 February. It then quickly emerged that a rising case load of respiratory diseases at Kirkland Life Care Center home for the elderly was a covid-19 outbreak. The respiratory problems had been

believed to be due to seasonal influenza or tuberculosis, which would have been normal for this time of the year. By the time it was identified as covid-19, the epidemic had already spread across the entire State and further into all of USA and Canada. This same mistake of believing a covid-19 to be an influenza outbreak is likely the reason why the epidemic remained undetected in Europe for three months.

The phylogenetic dendrogram provides additional information to the above case-tracing story. The 35-year old man must have infected at least two other travelers while being en-route in China, because there are two additional cases in Fujian and Hang Zhou, registered on 21 and 26 January, who share his exact same mutations. Apparently, he then infected three other persons either in Seattle airport or on the public transport from the airport. Two of these went on to travel to British Columbia in Canada, where their descendants were registered at the beginning of March. The third person went to Kirkland Life Care Center, and somehow mass-infected that elderly care home. As per the beginning of May, around half of all Washington State sequenced infections can genomically be traced back to Kirkland.

III a 3) The undetected European outbreak

Chart 5) shows the total dendrogram, but this time only with all European infections highlighted. Out of the 2,220 samples that are shown on the diagram, 403 are direct descendants from the Chinese outbreak. Plenty of Europeans became infected either in Wuhan directly or in other Chinese cities, which then in due course developed into various outbreaks all over Europe from the end of February onwards. The other 1,817 are NOT descendants from the Chinese outbreak, but originate on what is called the A2 clade of the epidemic. The A2 clade appears to have forked off from the others right at the first infection at the end of November, and has since become the source of the large majority of infections in Europe and North America. A2 should therefore be called the European clade.

The implication cannot be overstated. Since the A2 branch most probably originated at the end of November, all of its subsequent infection activity had nothing to do with the Wuhan and China outbreak. It was imported into Europe very early on, and created an European outbreak which unfolded entirely undetected by the European authorities. While the cameras and media attention zoomed into the Wuhan outbreak at the end of January and throughout February, there was a parallel and similarly sized outbreak unfolding somewhere in Europe which, however, went unnoticed, and to the day of writing this article is not recognized to have happened.

Authorities were on alert in both Europe and United States throughout February, for travelers from China or Asia with symptoms to be tested for covid-19. The result was a number of cases which were caught in January and February in various places, as can be seen on the Chinese clades. Among them was the earliest confirmed case in Europe, on 23 January in Paris, a 31-year-old woman who had just returned from a trip to Wuhan. But the authorities were oblivious to the much larger infection activity unfolding right under their eyes on European territory.

Chart 6) shows the same dendrogram, only restricted to the European subsample. The display algorithm arranges the clades in a different order, which is why the dendrogram looks different. It illustrates that 673 of the shown descendants originate from the China outbreak, and 2983 originate from the European outbreak. In terms of genetic diversity as displayed by the dendrogram, the European outbreak is more than four times larger than the Chinese one.

III a 4) Dynamics of the European outbreak

There are two problems with understanding the early dynamics of the undetected European outbreak. The first problem is that with one small but important exception, which will be explained further below, there are no European genetic mutations sequenced of the outbreak until the end of February. Not a single European health authority found a case on the A2 clade until then. The reason must be that authorities were not looking for one, because only travelers with a Chinese connection were tested. Therefore all infection activity must be indirectly inferred. The second problem is that those countries which have done a lot of sequencing since March are overrepresented in the sample, for instance the Belgians or the Dutch, whereas the Germans and the Italians, about whom we would like to know the most, have barely contributed useful data.

When sequencing did finally begin in Europe towards the end of February, the data points already represented a large genetic diversity with cases sitting on thousands of different clades. This made it difficult throughout March to use phylogenetics to draw conclusions, because the cases were scattered so widely, in genetic terms. Only towards the end of April did the picture become clearer of what might have happened ⁹.

Researchers from Los Alamos National Laboratory in the USA noted in their publication in late April that the particular mutation which is the root for the vast majority of European and American cases, called D614G on the A2 clade, appears to have higher evolutionary fitness for Europe than the mutations of the Chinese clades A1, A3 and B ¹⁰. However, the GISAID researchers doubt that such inferences can be drawn: "Changes in prevalence of these variants in the short term are expected and often driven by chance or so-called founder effects rather than immediate evolutionary pressure" ¹¹. The founder effect which the GISAID researchers suggest, would imply that the D614G is more prevalent, because its outbreak was larger. This can also be seen on Chart 7). The virus split into a Chinese and a European path already at the end of November. While the Chinese path was detected in December and triggered a response in January, the European path was not detected and triggered a response only in March. By this time, it was already several times larger than the Chinese outbreak. It was not larger because the D614G mutation is somehow better suited to Europeans, but because the Europeans did not notice the outbreak for three months.

Chart 8) shows that the United States also had two routes of infections. One route was to be infected directly in China, and the other route worked via imports from Europe from the D614G strain of mutations. On the North American subsample, 1,180 of its descendants came from the China outbreak, and 2,963 from the European outbreak. North America was attacked about three times stronger from Europe than from China.

As Chart 8) also shows, the US imported infections from Europe several hundreds of times, but there was one clade which particularly proliferated. Most other imports on the European clades tended to fade away. The same is true for the direct China imports, which did not spread as well from there into the rest of the United States. Apparently, just as D614G benefitted from a European founder effect, so did the Q57H mutation benefit from a North American founder effect, which appears to have been situated in New York City.

Chart 9) looks only at the American sample and selecting California and New York. With no surprise, the West Coast dominantly infected itself in China, and the East Coast dominantly in Europe.

Charts 10) and 11) show the distributions of the Spanish and French cases. Besides the Wuhan on the Chinese clade, Spain's infections are clustered around its own European clade. France is represented with two clusters on the dendrogram. One cluster is very close to the East Coast infection of USA and dominated by Ile de France (Paris), and the other cluster is dominated by cases from the French Alps.

Charts 12) and 13) show how the countries of Belgium, Netherlands and Austria are distributed across the entire European clade and do not cluster. The same could be said for Germany, however, since there are fewer sequences, it is less certain.

III b) More details on the proposal for how the epidemic spread in its early days in Europe

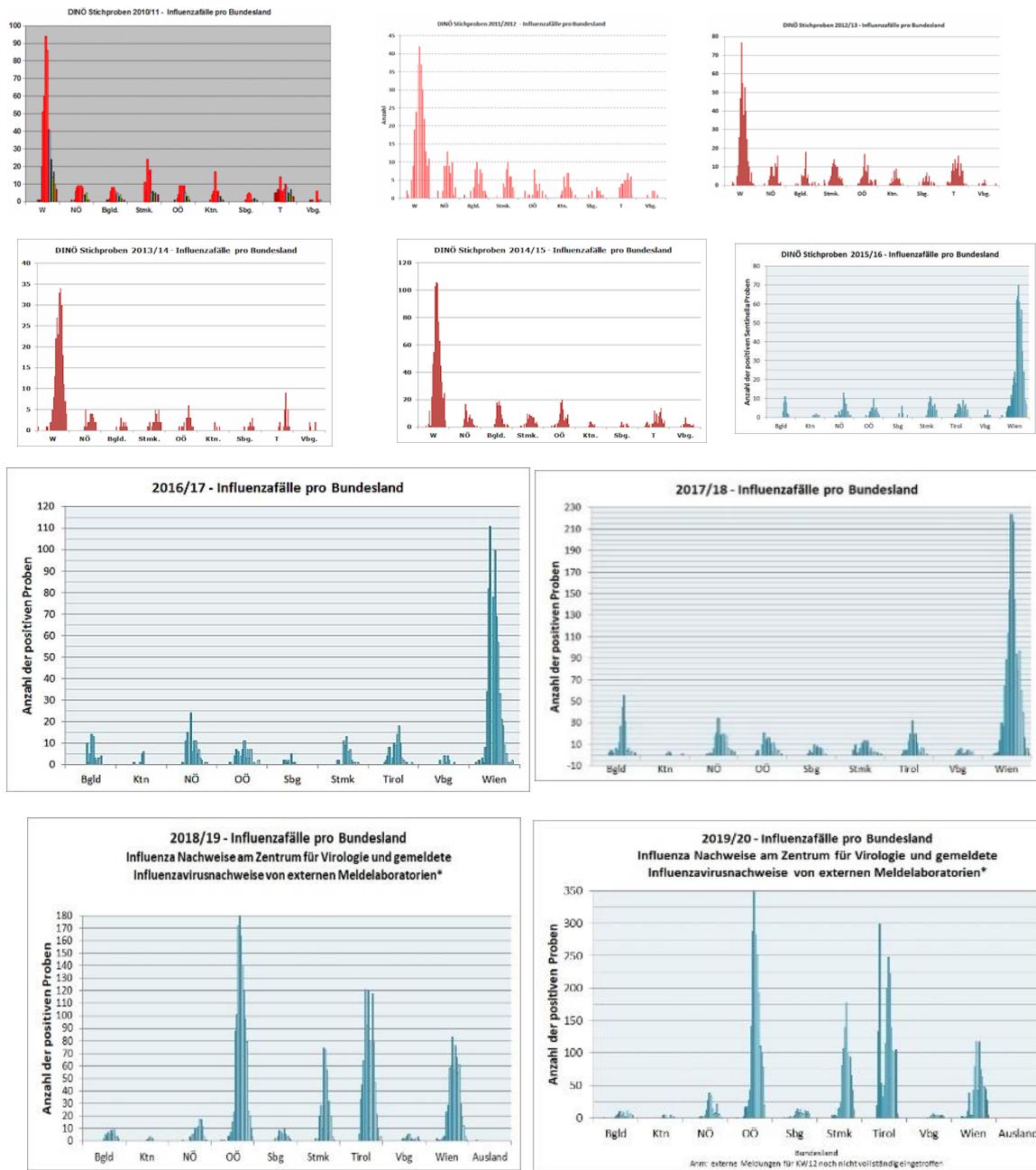
The following provides more detailed evidence to the main text on the proposed events of the early epidemic in Europe from December to February 2020. Understanding the evidence from the phylogenetic dendrogram will require reading the Section III a) of this Supplement as well.

III b 1) Stage 1: 01-06 December in Innsbruck, Tyrol, Austria

During CW 50 (09–15 December 2019) the alarm bells are ringing at the Austria influenza monitoring center at the Medical University of Vienna. It records an explosive jump in cases in the state of Tyrol, followed by another tripling of cases in CW 51.¹² Two elementary schools in the capital city of Innsbruck need to be closed because of the outbreak, in one of them 150 out of 250 students fell sick¹³. The monitoring center declares the opening of the influenza season and expects that the rest of Austria will soon follow.

There are several aspects that are unusual around this Tyrol outbreak. First, it arrives too early. Normally Tyrol begins its season towards the end of January. Second, its dimension: in Tyrol, there are ten times more positive influenza cases in CW 51 than at the peak in the strong influenza season of 2017/18. Third, in CW 52, the cases drop dramatically again (which however may be due to the doctors' offices being closed during the holidays). Fourth, only 5% of the suspected cases are actually influenza positive¹⁴. On 15 January, virologist Redburger-Fritz of MU Vienna calls this season "unusual" on national TV.¹⁵ Finally, against initial expectations, this influenza does not spread around the country. Throughout the season, only three states are strongly hit with cases, plus a little bit in the capital city of Vienna, but the other five states see almost no activity. The three states are Tyrol, Upper Austria and Styria. Also unusual: these are the same three states that were the only ones to have an influenza season in 2018/19, and with a similar influenza type A composition. In all previous years before that, influenza would strike Vienna heavily, and all other states equally mildly (Figure S III c).

Figure S III c) Austrian influenza patterns 2010 to 2020



Source: Medical University of Vienna, Influenza Monitoring System

Proposed explanation for what happened in Stage 1:

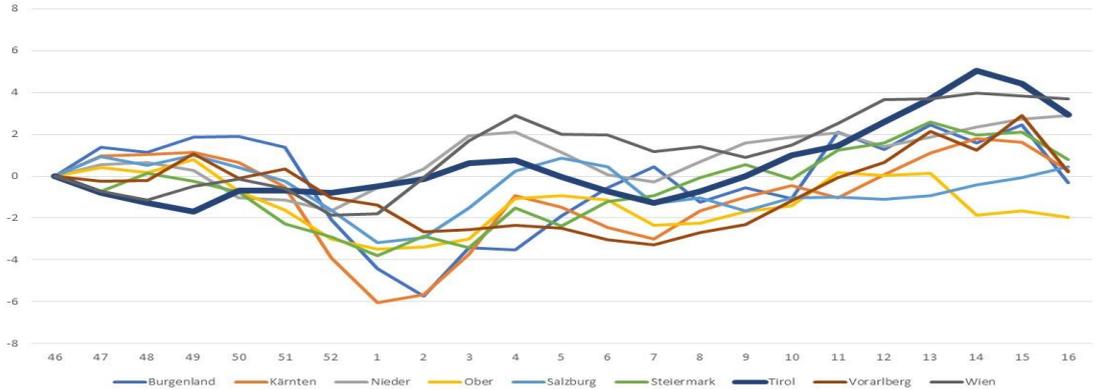
It is proposed, that the Tyrol outbreak in CW 50 and 51 was not an influenza outbreak, but a covid-19 outbreak. When patients went to the doctor they were tested for influenza. With comparatively low rates of positivity (the above mentioned 5%), virologists just found an echo of the previous season. It is not unusual for a particular virus pattern to perpetuate itself from season to season. Whatever was the reason for the strange influenza geographic pattern in 2018/19, it repeated itself in 2019/2020 in the same three states. Why the influenza pattern deviated so much in 2019 and 2020 from all previous years since 2010, as is shown in Figure S III c) also deserves an epidemiological answer. Whether this is related in any way to covid-

19 cannot be established with the current data. What is certain, is that the Tyrol outbreak in December 2019 has no precedence anywhere in Austria over the previous ten years. Note also that the units of the y-axis are twice as high in 2020 as in all previous years. Therefore it is proposed, that influenza was registered so massively and so early in Tyrol only because patients went to the doctors for a different and parallel kind of respiratory disease. However, this covid-19 disease was unknown at the time and therefore was not registered.

The affected patients must have spent an uncomfortable holiday season with what they thought was an unusually stubborn cold. However, it was not a cold. Two weeks later in CW 1 and 2, there are 1.7 persons per 100,000 inhabitants additional deaths in Tyrol compared to the national average. The reason could not have been influenza. The outbreak of the parallel occurring ordinary influenza in the states of Upper Austria and Styria was not unusually fatal. There is no above average mortality in these two states versus the national average. So the additional deaths in Tyrol cannot be ascribed to actual influenza prevalence.

Figure S III d) Austria excess mortality

Unit of y-axis is # of persons per 100,000 population. The horizontal 0 line represents the average expected mortality on the basis of three previous years 2016, 2017 and 2019 per calendar week. The lines per state represent excess (above the 0 line) or deficit mortality (below the 0 line) versus their own respective mortalities in those previous years. The lines are cumulative with base 0 in week 46 in 2019.



Source: Own calculations based on mortality statistics provided by Austria Statistics

The 1.7 number translates into 27 additional deaths during those two weeks. There were 275 people who died during these two weeks in Tyrol, 245 of them above the age of 65. That is a normal ratio for January, in line with previous years. Also, the mortality ratio itself is not unusual, the accumulated total even stays slightly below the expected line. Those 26 additional deaths arise only from the fact, that the rest of the country experienced 14% fewer deaths compared to previous years, while Tyrol only experienced 4% less. No current observer would have suspected anything unusual. It was only unusual in hindsight compared to the national Austrian average.

When we include CW 52 and CW 3, the additional deaths versus the national average in those four weeks amount to 45 persons in Tirol. By CW 5, Tirol reverts back to the national average. It only rises above the national average again in CW 10, (02–08 March) as a result of the reimported second wave of covid-19 in March. If the IFR from the Gangelt case study and the U.S. CDC is applied, then those additional 45 deaths represent around 11,000 infections which may have been circulating around Tyrol from the middle to end of December. This estimate has a very large range of uncertainty, as there might have been a number of other local known,

unknown or stochastic factors, which influenced the mortality rate in Tyrol. But even incorporating this range of uncertainty, something was happening in Tyrol in December, and it was not classic influenza.

The authorities say that they found 450 influenza positive tests in Tyrol in CW 50 and 51, which they estimated to have been 5% of the total reported patients, which makes for a total number of around 8,550 infections with an unknown disease. This number roughly matches with the 11,000 estimate from above.

As to who would have brought covid-19 into Tyrol can only be guessed at this point. There must have been a super-spreader event, because CW 50 is rather close to where the A2 clade forked off, in a place probably in China. There was not much time that could have been spent on a local infection cascade in Tyrol.

We know that two elementary schools needed to be closed, one in Innsbruck Igls on 13 December and one in Innsbruck Angergasse on 16 December. This brings some Santa Claus events into focus. Santa Claus arrived in the mountain village of Igls on Sunday 01 December, on 05 December in Innsbruck city center, and on 06 December on the *Christkindlmarkt* in the Hungerburg mountain above Innsbruck¹⁶. If one or several of the Santas and angels at these events were infected with covid-19, these events could have easily infected hundreds of elementary school children, which could have subsequently led to school closures two infection cycles later. It is well understood that children can also be infected with covid-19.¹⁷

The hypothesis can be tested in two ways. If it is correct that the proposed Tyrol covid-19 outbreak was concentrated on Innsbruck, then those additional 45 deaths should be concentrated in Innsbruck as well. In normal years, the month of January sees around 105 deaths in Innsbruck. Given the low national average, it should have had only around 90 deaths in January 2020 (105*86%). If the proposal is right, then there should be 45*75% (three weeks in January, one week in December) = 34 additional deaths in January, for a total of 124 deaths in Innsbruck in January, plus/minus a considerable range of uncertainty. As of writing this article, the official mortality statistics are not yet available for the local registries, and their processing has been curtailed by the covid-19 emergency in Tyrol in March and April. However, if a special investigation would be undertaken, it could either substantiate or refute the proposal. However, it is possible to note that there were unusually many public obituaries in newspapers and internet pages in Innsbruck in January 2020.

A second test would be to conduct antibody tests on the children of both the schools of Igls and Angergasse, and on the Santa Clauses and accompanying angels that were running those events. If they test positive, then tracing back the travel histories of those Santas might lead to the real European patient 0. If the schools do not test positive, then there might have been some other events or occasions where Innsbruck or Tyrol got infected – or indeed the whole Tyrol hypothesis might not hold up. It will require more primary data to either confirm or reject the proposal.

In the beginning of March, the Tyrol winter vacation hotspot of Ischgl became known and was confirmed by numerous contact tracings all around Europe, as being a super-spreader center of the European covid-19 epidemic. A throng of 5400 vacationers are known to have been infected there, two thirds of them from Germany¹⁸. The phylogenetic dendrogram of the mutations shows without any doubt that it would be pure coincidence, for Tyrol to become a center of the outbreak twice. If the putative Innsbruck outbreak had propagated to Ischgl during

January, then there would be direct descendants of that early branch visible all over Europe. But there are only very few of them, which were later found in the Austrian city of Graz.

The story could have ended in Innsbruck and we might never have known about it, except that there must have been infection chains that found their way to Munich and from there to the world, which is the story of the next stages. There could not have been direct descendants from Innsbruck in the ski resorts. Those were re-imports from Europe several weeks later.

III b 2) Stage 2: 03-04 January at the Four Hills Tournament in Innsbruck

If there was a Tyrol covid-19 outbreak (in Innsbruck or elsewhere), then it had almost faded away on its own without leaving any known direct descendants on the phylogenetic dendrogram, and without the authorities having noticed anything about it. In December it would have been impossible to think of a corona-virus epidemic, because the events in Wuhan were still unknown at the time. Even in retrospect, it does not seem to have been considered yet, that the highly irregular influenza season in Innsbruck could have been a covid-19 outbreak.

But chart 20) on the phylogenetic dendrogram shows that there must have been one event in CW 1, where several infection chains were generated at once, which then independently led to the three super-spreading events which likely occurred in Munich and in New York one month later, as described in Section 4 of the main text. In addition, we know for sure that all other Austrian states except Tyrol, and also uniquely Bavaria in Germany, witnessed a rise in their mortality against the expected trend from CW 3 onwards, lasting until about CW 5, when reverting back to the trend of the previous years (Figures 2 and 12 in the main report). Those three weeks would fit perfectly with a holiday period infection in Innsbruck. So we are searching for an event with the capacity to create multiple independent infection chains, in calendar week 1, which was attended by many Austrians and Bavarians, but Austrians and Bavarians only, because only they had irregular mortality patterns from CW 3 onwards.

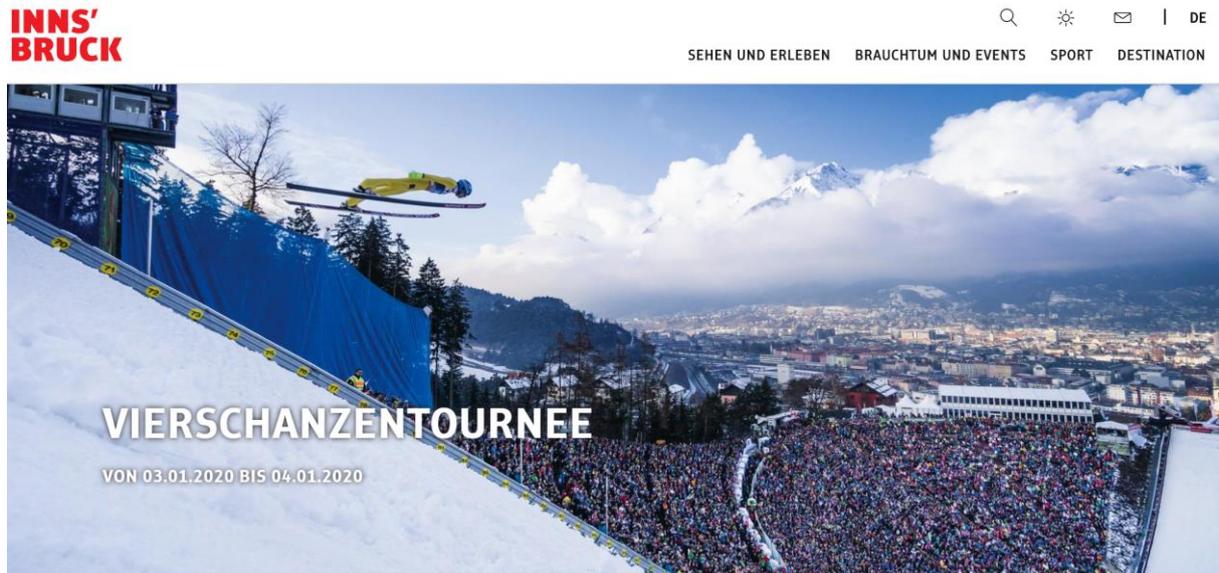
Proposed explanation for what happened in Stage 2:

Christmas holidays are peak season in Tyrol. The state with its 760,000 inhabitants, provided 11.2 million hotel nights for 2.5 million guest arrivals between November 2019 and January 2020. Of all the guests, 56% came from Germany, 11% from Netherlands, 5% from Switzerland and 7% from Austria¹⁹. Residents from Bavaria are much overrepresented among the Germans because of the geographic proximity. If there was a covid-19 infection hotspot in Innsbruck or close to Innsbruck, then tourists from all over Europe could have infected themselves there – and possibly did. But this was not yet on the dimension of Ischgl at the beginning of March. It was a relatively small local outbreak in a city of 133,000 inhabitants, of whom around 10,000 may have been infected at some point, many of them children, if the proposal is true. A hotel or restaurant bar could not have generated enough infection chains to cause a noticeable outbreak among the roughly 20 million inhabitants of Bavaria and Austria. Moreover, it needed to be an event which infected almost only Bavarians and Austrians and none or few other Europeans, with the exception of possibly Polish visitors from Silesia.

The most plausible event where this super-spreading could have occurred is the Four Hills Tournament of ski-jumping, which is conducted each year in the four Alpine cities of Oberstdorf, Garmisch-Partenkirchen, Innsbruck and Bischofshofen. Each of the four locations

receives around 30,000 visitors for their two days of activities. The tournament is the annual highlight of the ski-jumping calendar and an important regional celebrity event for South Bavarians and Austrians. Innsbruck received the tournament on 03/04 January.

Figure S III e) Impression of Innsbruck Four Hills Tournament site and spectators



Source: Innsbruck Info

III b 3) Stage 3: 28 January in the Munich region of Bavaria

In the course of January there is a similar influenza pattern in Upper Bavaria (the region south of Munich and adjacent to the Alps) as in Tyrol in December. The influenza season started too early and it arrived in multiple strength over the previous years, before suddenly subsiding again. However, only about one third of those with symptoms test positive for influenza. Two elementary schools need to be closed in Starnberg, a regional hub city south of Munich. It is unclear why mostly Upper Bavaria was hit with influenza and that it did not spread further²⁰.

In the same region a few kilometers south of Munich, Germany becomes aware of its first domestic covid-19 case on 28 January. A 33-year old manager tests positive in what became known as the Webasto outbreak, where ultimately 16 persons became ill. This Webasto outbreak has been covered intensively by the authorities and science, because it was also considered to be the first case where transmission occurred outside of China.

Proposed explanation for what happened in Stage 3:

It is proposed that the regional influenza outbreak in the region south of Munich in January was actually a covid-19 outbreak. We have one small glimpse into how one of its likely hundreds of infection chains in the region took course. The only reason we know about it, is because a Chinese participant of a business meeting returned to her hometown in Shanghai and received a positive PCR test there. If it had not been for this coincidence, the Webasto outbreak would most likely have never been detected. Other infections on the mutation clade of the Webasto

outbreak, were ultimately discovered in March in Belgium, Finland, Brazil, Austria and Guangzhou (Chart 14).

An unbiased evaluation of the primary data would conclude that the Chinese business woman caught the infection in Bavaria and not the other way around. This will be elaborated in more detail in the following.

There is no direct evidence for the nationality of the person who infected himself at the end of November and then became the founder of the European A2 branch characterized by the D641G mutation (see Section S III A 4). It is generally assumed that it was a Chinese person, but there is no direct evidence of this. It could just as well have been a European person who collected the virus at the end of November, either in China or any other place, and who then started the European infection chain in Europe in the beginning of December, possibly in Innsbruck as explained above. Besides the Webasto outbreak, there are various pieces of evidence for this proposal.

First, we know of at least one further international infection chain which links to the earliest fork in late November, this being a Thai infection chain which was detected on 29 January.

Second, there are only a handful of Chinese cases on this early A2 clade, and none on the forks after mid-January. If this clade had high prevalence in China until the end of January or even early February as is currently assumed, then there should have been many more Chinese cases being detected along all forks of this clade. It seems more likely that the few cases on the early A2 clade are re-imports into China, by Chinese who infected themselves in Europe. Clades which forked off early from the A2 clade have been found on all continents, so there should be nothing surprising of reimported infections into China from a European source. Another possibility is that those Chinese cases are wrongly assigned to the A2 clade, given the high degree of statistical uncertainty on the very early cases. One of the Wuhan infections which the algorithm assigns to have forked off after 28 December carries the D614D mutation, even though previous forks already carry the D614G mutation. This would indicate a wrong assignment.

The third piece of evidence is a clade which forked off the A2 on 28 December 2019 (corresponding to an infection in mid-December), on which many members were identified in Graz in Austria in late March and the beginning of April. Importantly, these Austrian members had already acquired the D614G mutation. This could be evidence that as of mid-December the A2 clade was already residing in Austria (Chart 15).

The Webasto case study in detail

The most important piece of evidence that Bavaria was the source of infection of the Webasto outbreak lies in the case details itself. On 28 January there was an announcement that the first German national had been infected during a business meeting on 20 or 21 January at headquarters of his employer Webasto. The 33-year old man had only mild symptoms and stayed home for three days from 24 to 26 January²¹. He would normally not have considered visiting a doctor²². The reason why he got himself tested for covid-19, is because there was a Chinese colleague who attended the seminar as well. After returning to Shanghai where she lived, she developed symptoms on 23 January and tested positive on 27 January. When this message was relayed back to Webasto, the 33-year old decided to arrange a test. When his test was confirmed positive on 28 January, the authorities and the Webasto management took

swift action to close the company and contact-trace everybody. Ultimately 16 people tested positive in the Webasto context.

Two academic journal articles were published on the case, one very early already in January²³ and one in May with more distance to the event, but with more in-depth investigations including genetic analysis of the German cases and detailed study of the electronic calendars²⁴. The two accounts partially contradict each other, and to some degree they both needed to rely on memories of the persons involved, which may suffer from witness bias. They also numbered the patients differently. Finally, there is missing clarity on the Chinese side of the story. The Chinese business woman said that she had been visited by her parents from Wuhan region three days before she left for Germany. But there are contradictory messages about whether her parents had been confirmed covid-19 positive later on. The May article cites the January article as a reference for that matter, but the January article is actually silent on it.

It is usually and widely assumed, if not asserted, that the 33-year old Webasto employee received the infection from the Chinese business women, which seems to make perfect sense in the context of those times. He sat next to her for one hour on 20 January in a small 12 sqm meeting room with four persons present. However, none of the investigations consider that it just as well could have been the other way around, indeed even more likely so. The Chinese business woman had no serious symptoms while being in Germany, and she did not seem to have infected any other employee at Webasto. Her movements were meticulously retraced, and nobody in her path, not at the hotel, nor in the taxi, nor the air plane had symptoms or tested positive.

There was a second positive Webasto patient who does not recall any encounter with the Chinese business woman. He did have meetings with the first patient. The researchers cannot explain how patient 2 could have been infected at all if the Chinese business woman was the source. The third positive person at Webasto also had no interaction at all with the Chinese woman, but did have contacts with the first patient on 20, 21 and 23 January. The researchers believe that patient 3 became infected by patient 1, which assumes that both of them had incubation periods of only two days each, which speaks very much against the literature^{25 26 27}²⁸. A fourth patient did have encounters with the Chinese business women, but only remotely. He also had several contacts with patient 1 (Figure S III f).

Without further data and taking an unbiased view, the chain of events makes it far more likely that the assumed German patient 1 is actually patient 0, and that he was the one who infected his Chinese colleague, as well the three other German colleagues. While genetic sequencing analysis was performed on all German cases, which prove that they were all linked, no sequencing was done on the Chinese specimen, and especially not the parents of the Chinese business woman.

Both the authorities and the researchers assumed that because the so-called patient 1 had not been abroad for the previous 14 days, that he could not have been infected other than from the Chinese patient. The possibility that he could have infected himself in Bavaria was at no point considered. Maybe it can be revisited whether he had been in contact with persons who visited the Four Hills Tournament in Innsbruck and became ill afterwards.

Figure S III f) Webasto case tracing chart

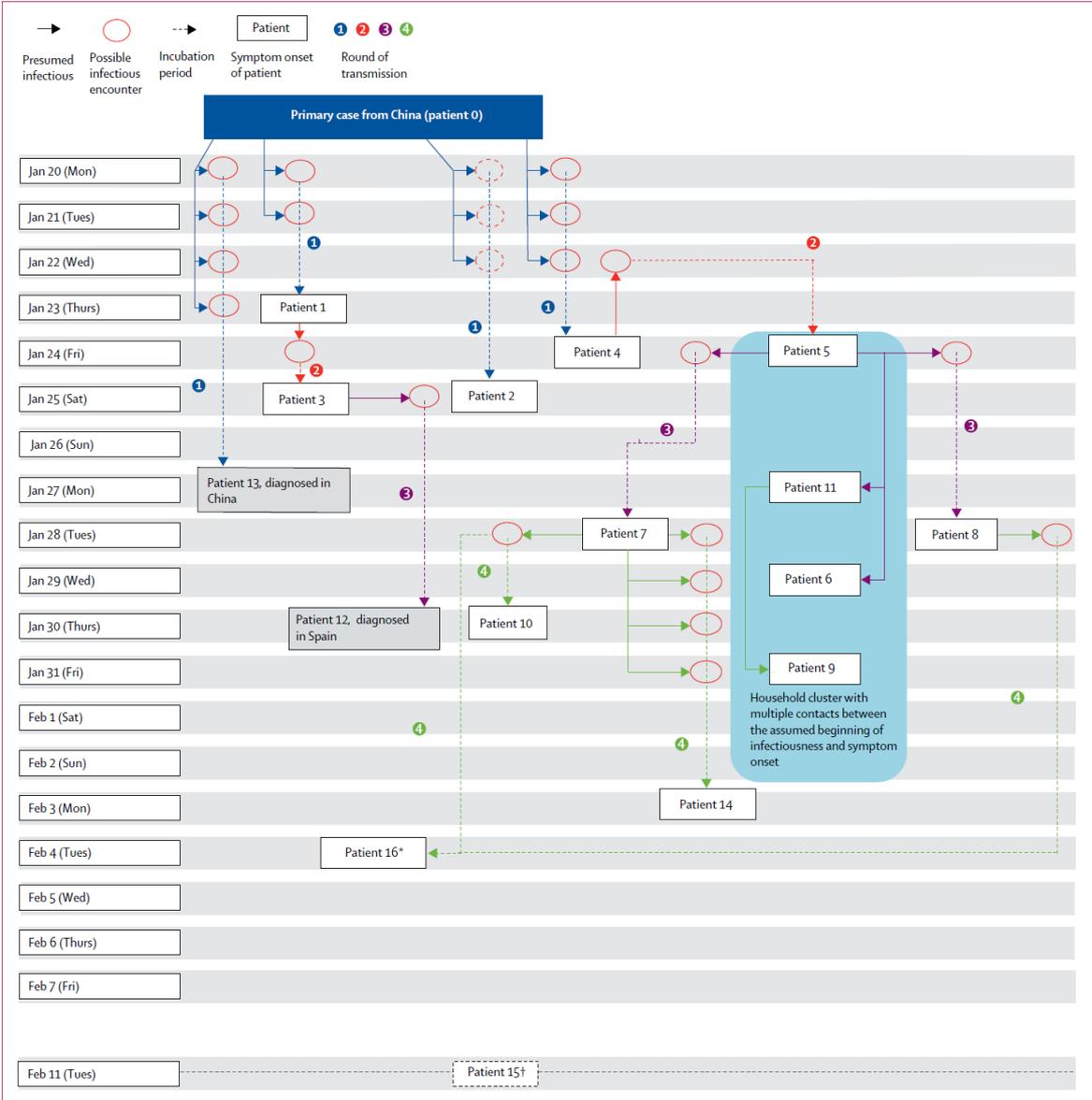


Figure: Transmission chain of COVID-19 satellite outbreak in Bavaria, Germany, in January–February, 2020. Boxes denote the day of symptom onset of cases, transmission rounds (arrows) are numbered and displayed in different colours. Red circles indicate the encounter when transmission is likely to have occurred; transmission from patient 0 to patient 2 is confirmed by whole genome sequencing, but no specific encounter could be identified. Potential presymptomatic infectious encounters are only included if no other encounter could be identified. Dotted arrows indicate the incubation period (transmission event until presentation of first symptoms), solid arrows lead from source cases to likely infectious encounters with recipient cases. For cases, the infectious period was assumed to start 2 days before symptom onset. SARS-CoV-2=severe acute respiratory syndrome coronavirus 2. *Also met patient 8 on Jan 28, but transmission was more likely through patient 7. †Asymptomatic household contact of case 2, with contact assumed Jan 25–28. Tested positive for SARS-CoV-2 on Feb 11.

Source: Böhmer et al, Lancet Infectious Diseases published online May 2020

A prominently published paper in the *Journal of Medical Virology* has been substantiating the Bavaria proposal already since the beginning of March. Zehender and Galli from the university research group and infectious diseases department of Milano’s Sacco Hospital had identified the evidence that early Italian covid-19 cases must have had a common ancestral origin in Munich²⁹. Their proposal was discarded because the genetic distance between the Webasto case and the Italian cases was too great for the Italians to be direct descendants of the Webasto outbreak. On the other hand, it has been considered certain that the Webasto case originated in China, and could not have had Munich-based ancestors. So Zehender and Galli’s findings have been ignored (Chart 16).

III b 4) Stage 4: 26-29 January ISPO sports business professional trade fair

On the phylogenetic dendrogram, one of the three early European clades leads to what must have been a large-scale super-spreader event in CW 5, where the virus was spread simultaneously to dozens of countries on all continents around the world. Chart 17) in Supplement IV shows this part of the dendrogram. It looks different from the other super-spreader events because it does not have a straight vertical line, but instead shows several initial cascading forks. However, each of these forks have the same genetic distance (3,968 units), which means that there were many infected persons at this super-spreading event who started off infection chains (at the other super-spreader events like in Kirkland Life Care Center, it was only one person who was a super-spreader).

The earliest known registration on this clade is the first Italian case in Codogno, who Zehender and Galli linked to a common ancestor with the Webasto cases.

Proposed explanation for what happened in Stage 4:

From 26-29 February, Munich hosted the ISPO trade fair. It is advertised as: “The largest trade fair for sports business. At ISPO Munich the entire sports world will gather again to shape the future in the segments Snowsports, Outdoor, Health & Fitness, Urban, and Teamsports. Look forward to four days with over 2,800 exhibitors and more than 80,000 visitors”³⁰.

ISPO would fit the phylogenetic information exactly. It happened at the right time, and it would have been an event where tens of thousands of participants from countries around the world were present. They could have carried the virus from Munich to the world.

It may be entirely coincidental, but it is conspicuous that there is a possible link between: 1) a sports business and sports textile trade fair in Munich which fits exactly the phylogenetic information, and 2) a Polish sports wear fashion label 4F who is the main sponsor of the preceding assumed super-spreader event at the Four Hills Tournament in Innsbruck, and 3) the first Italian confirmed case Mattia, who is known to be very active in recreational sports, and for whom it could not be identified how and where he became infected. Maybe additional case tracing work can establish a link between Mattia and professional visitors at ISPO in Munich.

III b 5) Stage 5: 04-06 February Munich Fabric Start 2020 professional trade fair

The second early European branch leads to what must have been another large-scale super-spreader event in CW 6, where the virus was again spread simultaneously to dozens of countries on all continents around the world. Chart 18) in Supplement IV shows this part of the dendrogram. The earliest known registration on this G50N mutation clade is in England on 23 February.

Proposed explanation for what happened in Stage 5:

From 04 February to 06 February the Munich Fabric Start Trade Fair was held in Munich, and it is one of the key events in the global fabrics industry. The fair is advertised as: “About 1,000 exhibitors from 40 countries present some 1,800 collections and a comprehensive range of

fabrics, trimmings and additional, fabric finishing, full-packaging services and manufacturing, thus ensuring the competence and professionalism required by more than 20,350 trade visitors for a perfect season opening”³¹.

It is proposed that one of the infection chains which had been circulating in the Upper Bavaria region during January was also prevalent at the MFS fair. It would have needed only one prominent exhibitor to have been highly infectious during those two days – most certainly a non-symptomatic index person – who could have infected scores of professional visitors from all corners. The infection could have been spread via close encounter talking, or via smear infection on the fabrics. The other, equally likely transmission path, could have been that one highly infectious visitor touched many of the fabrics of the fair, which were then shortly afterwards touched by other visitors. It is in the nature of a fabric fair, that the fabrics are sensed and touched. The promotional pictures of the fair on the internet show almost exclusively visitors who are touching and feeling fabrics.

One of the main fabric manufacturing centers in Italy is the Seriano Valley, north of the city of Bergamo. At the entrance to the valley is the city of Alzano. Internal documents cited by Italian newspapers indicate that the Alzano hospital handled a number of serious pneumonia cases as early as 12 February, which were likely to have been covid-19³². Three weeks later, Bergamo’s hospital system began to collapse under the surge of critically ill patients.

III b 6) Stage 6: 03-12 February Men’s and Women’s Fashion Week in New York

The third European clade on the phylogenetic dendrogram has a cluster where most of the cases of both Paris and New York are simultaneously situated. Chart 19) zooms into this clade with the Q57H mutation. Here are now shown the cases found in New York State, as well as Ile de France and Haut de France, which is Paris and the region north of Paris. Around 90% of the Paris and Northern French sites are in the same cluster as New York and the East Coast. Only a joint super-spreading event could have created such an outcome.

Proposed explanation for what happened in Stage 6:

The proposed event where this could have happened was the New York Fashion Week first for Men and then for Women, respectively on 03–05 February and 06–12 February³³. That the other European fashion capital of Milano is not prevalent with Italian cases in this clade of the dendrogram indicates that the super-spreading was restricted to a French fashion participation cluster. Indeed, it is almost impossible to think of any other event that could have achieved the peculiar outcome of the phylogenetic dendrogram. The virus would have continued to circulate in both cities among the fashion industry for a few more weeks, hidden behind a façade of an industry that is socially not allowed to become ill. By the time the outbreak was fully noticed in March, Paris and New York became two of the heaviest attacked cities in the world.

The summary of all proposed events is shown on Chart 20) in Supplement IV

¹ <https://www.gisaid.org/references/statements-clarifications/different-types-or-clades-of-the-virus-and-their-origin/#c501>

² <https://www.economist.com/science-and-technology/2020/05/02/the-pieces-of-the-puzzle-of-covid-19s-origin-are-coming-to-light>

³ Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus–Infected Pneumonia: Qun Li, M.Med., Xuhua Guan, Ph.D., Peng Wu, Ph.D., Xiaoye Wang, M.P.H., Lei Zhou, M.Med., Yeqing Tong, Ph.D., Ruiqi Ren, M.Med., Kathy S.M. Leung, Ph.D., Eric H.Y. Lau, Ph.D., Jessica Y. Wong, Ph.D., Xuesen Xing, Ph.D., Nijuan Xiang, M.Med., et al. N Engl J Med 2020; 382:1199-1207 DOI: 10.1056/NEJMoa2001316
<https://www.nejm.org/doi/full/10.1056/nejmoa2001316?sort=oldest&page=0>

⁴ <https://www.caixinglobal.com/2020-02-23/another-study-claims-wuhan-seafood-market-may-not-be-source-of-covid-19-outbreak-101519302.html>

⁵ <https://nypost.com/2020/03/27/shrimp-vendor-at-wuhan-market-may-be-coronavirus-patient-zero/>

⁶ <https://www.economist.com/science-and-technology/2020/05/02/the-pieces-of-the-puzzle-of-covid-19s-origin-are-coming-to-light>

⁷ <https://www.seattletimes.com/seattle-news/times-watchdog/lost-battle-how-coronavirus-overwhelmed-washington-states-efforts-to-contain-it/>

⁸ First Case of 2019 Novel Coronavirus in the United States by Michelle L. Holshue, M.P.H.
N Engl J Med 2020; 382:929-936
<https://www.nejm.org/doi/full/10.1056/NEJMoa2001191>

⁹ <https://nextstrain.org/narratives/ncov/sit-rep/2020-05-15?n=2>

¹⁰ Spike mutation pipeline reveals the emergence of a more transmissible form of SARS-CoV-2
B Korber, WM Fischer, S Gnanakaran, H Yoon, J Theiler, W Abfalterer, B Foley, EE Giorgi, T Bhattacharya, MD Parker, DG Partridge, CM Evans, TM Freeman, TI de Silva, on behalf of the Sheffield COVID-19 Genomics Group, CC LaBranche, DC Montefiori; bioRxiv 2020.04.29.069054;
doi: <https://doi.org/10.1101/2020.04.29.069054>

¹¹ <https://www.gisaid.org/references/statements-clarifications/different-types-or-clades-of-the-virus-and-their-origin/#c501>

¹² <https://www.virologie.meduniwien.ac.at/wissenschaft-forschung/virus-epidemiologie/influenza-projekt-diagnostisches-influenzanezwerk-oesterreich-dinoe/aktuelle-saison-20192020/>

¹³ <https://www.tt.com/artikel/16452914/150-kinder-krank-grippewelle-legt-schule-in-innsbruck-lahm>

¹⁴ <https://www.tt.com/artikel/15372594/grippewelle-in-tirol-das-sind-die-besonderheiten-dieser-saison>

¹⁵ <https://orf.at/stories/3151048/>

¹⁶ <https://www.christkindmarkt.cc/de/highlights-zur-adventzeit/nikolausumz%C3%BCge/90-0.html>

¹⁷ <https://www.nature.com/articles/d41586-020-01354-0>

¹⁸ <https://www.suedkurier.de/ueberregional/panorama/die-akte-ischgl-1000-seiten-polizeibericht-5380-krank-ur-lauber-25-tote;art409965,10508208>

-
- ¹⁹ <https://www.ttr.tirol/statistik/statistik/wintersaison-201920>
- ²⁰ <https://www.sueddeutsche.de/bayern/bayern-grippe-influenza-zahlen-1.4772104>
- ²¹ <https://www.zeit.de/wissen/umwelt/2020-01/coronavirus-webasto-ausbruch-bayern-china-ansteckung-deutschland>
- ²² <https://www.br.de/nachrichten/bayern/coronavirus-patient-nummer-1-wie-ich-die-quarantaene-erlebte,Rrm4UI8>
- ²³ Transmission of 2019-nCoV Infection from an Asymptomatic Contact in Germany, by Rothe et al, in Letters to the New England Journal of Medicine, 30 January 2020
<https://www.nejm.org/doi/10.1056/NEJMc2001468>
- ²⁴ Investigation of a COVID-19 outbreak in Germany resulting from a single travel-associated primary case: a case series, by Merle M Böhmer, Udo Buchholz, Victor M Corman, Martin Hoch, Katharina Katz, Durdica Marosevic, et al. in The Lancet Infectious Diseases Journal on May 15, 2020
[https://doi.org/10.1016/S1473-3099\(20\)30314-5](https://doi.org/10.1016/S1473-3099(20)30314-5)
- ²⁵ Serial interval of novel coronavirus (COVID-19) infections: Hiroshi Nishiur, Natalie M., Lintona Andrei, R.Akhmetzhanova
<https://doi.org/10.1016/j.ijid.2020.02.060>
- ²⁶ Estimating the generation interval for COVID-19 based on symptom onset data: Tapiwa Ganyani, Cecile Kremer, Dongxuan Chen, Andrea Torneri, Christel Faes, Jacco Wallinga, Niel Hens
<https://doi.org/10.1101/2020.03.05.20031815>
- ²⁷ McAloon, C. G. (2020). The incubation period of COVID-19 – A rapid systematic review and meta-analysis of observational research. medRxiv preprint doi: <https://doi.org/10.1101/2020.04.24.20073957>.
<https://www.medrxiv.org/content/10.1101/2020.04.24.20073957v1>
- ²⁸ Quantifying SARS-CoV-2 transmission suggests epidemic control with digital contact tracing
Luca Ferretti, Chris Wymant, Michelle Kendall, Lele Zhao, Anel Nurtay, Lucie Abeler-Dörner, Michael Parker, David Bonsall, Christophe Fraser; 31 March 2020 on Science First Release
DOI: 10.1126/science.abb6936
<https://science.sciencemag.org/content/368/6491/eabb6936>
- ²⁹ Zehender, G, Lai, A, Bergna, A, et al. Genomic characterization and phylogenetic analysis of SARS-COV-2 in Italy. J Med Virol. 2020; 1– 4.
<https://doi.org/10.1002/jmv.25794>
- ³⁰ <https://www.ispo.com/en/munich>
- ³¹ https://www.munichfabricstart.com/tl_files/mfs/downloads/MFS_SalesKit_en.pdf
- ³² <https://apnews.com/de2794327607a3a67ed551f0b6b71404>
- ³³ <http://www.modemonline.com/fashion/fashion-weeks/fall-winter-2020>