Accelerated Article Preview

Cryptic transmission of SARS-CoV-2 and the first COVID-19 wave

Jessica T. Davis, Matteo Chinazzi, Nicola Perra, Kunpeng Mu, Ana Pastore y Piontti, Marco Ajelli, Natalie E. Dean, Corrado Gioannini, Maria Litvinova, Stefano Merler, Luca Rossi, Kaiyuan Sun, Xinyue Xiong, Ira M. Longini Jr, M. Elizabeth Halloran, Cécile Viboud & Alessandro Vespignani

This is a PDF file of a peer-reviewed paper that has been accepted for publication. Although unedited, the content has been subjected to preliminary formatting. Nature is providing this early version of the typeset paper as a service to our authors and readers. The text and figures will undergo copyediting and a proof review before the paper is published in its final form. Please note that during the production process errors may be discovered which could affect the content, and all legal disclaimers apply.

Article

Cryptic transmission of SARS-CoV-2 and the first COVID-19 wave

Jessica T. Davis[1,11], Matteo Chinazzi[1,11], Nicola Perra[1,2,11], Kunpeng Mu[1], Ana Pastore y Piontti[1], Marco Ajelli[1,3], Natalie E. Dean[4], Corrado Gioannini[5], Maria Litvinova[3], Stefano Merler[6], Luca Rossi[5], Kaiyuan Sun[7], Xinyue Xiong[1], Ira M. Longini Jr[8], M. Elizabeth Halloran[9,10], Cécile Viboud[7] & Alessandro Vespignani[1✉]

Considerable uncertainty surrounds the timeline of introductions and onsets of local transmission of SARS-CoV-2 globally[1–7]. Although a limited number of SARS-CoV-2 introductions were reported in January and February 2020[8,9], the narrowness of the initial testing criteria, combined with a slow growth in testing capacity and porous travel screening[10], left many countries vulnerable to unmitigated, cryptic transmission. Here we use a global metapopulation epidemic model to provide a mechanistic understanding of the early dispersal of infections, and the temporal windows of the introduction and onset of SARS-CoV-2 local transmission in Europe and the United States. We find that community transmission of SARS-CoV-2 was likely in several areas of Europe and the United States by January 2020, and estimate that by early March, only 1 to 3 in 100 SARS-CoV-2 infections were detected by surveillance systems. The modelling results highlight international travel as the key driver of the introduction of SARS-CoV-2 with possible introductions and transmission events as early as December 2019–January 2020. We find a heterogeneous, geographic distribution of cumulative infection attack rates by 4 July 2020, ranging from 0.78%–15.2% across US states and 0.19%–13.2% in European countries. Our approach complements phylogenetic analyses and other surveillance approaches and provides insights that can be used to design innovative, model-driven surveillance systems that guide enhanced testing and response strategies.

A few weeks after the initial announcement of a cluster of atypical pneumonia cases in Wuhan, China, the first confirmed cases of COVID-19 in the United States (US) and Europe were detected on January 21, 2020 in Washington state[1] and on January 24, 2020 in France[2]. Although many more states and countries began reporting initial introductions in the following weeks, only a few cases were detected daily during this time period (see Fig. 1a), and most countries adopted a testing policy that targeted symptomatic individuals with a travel history linked to China. Several reports suggest the introduction of SARS-CoV-2 occurred earlier than initially recognized[3–8], raising questions about the effectiveness of the initial testing policies and travel-related restrictions, as well as the extent to which the SARS-CoV-2 virus spread through cryptic transmission in January and February 2020. In order to address these questions, we use the Global Epidemic and Mobility Model (GLEAM), a data-driven, stochastic, spatial, and age-structured, metapopulation epidemic model[11,12], to study the global dynamic underlying the evolution of the COVID-19 pandemic in Europe and the United States (US). Our model maps the plausible pathways of the pandemic using information available at the early stages of the outbreak and provides

a global picture of the cryptic phase as well as the ensuing first wave of the COVID-19 pandemic.

We consider data concerning the continental US and 30 European countries (the full list is reported in Extended Data Table 1). The model integrates real time human mobility and population data with a mechanistic epidemic model at a global scale, incorporating changes in contact patterns and mobility according to the non-pharmaceutical interventions (NPIs) implemented in each region. It is calibrated on international case introductions out of mainland China at the early stage of the pandemic using an Approximate Bayesian Computation (ABC) methodology[13]. The model returns an ensemble of stochastic realizations of the global epidemic spread including international and domestic infection importations, incidence of infections, and deaths at a daily resolution (see the Methods section for details). In the following text we provide a detailed discussion of the analyses and results concerning European countries and the US states, however, to further test and validate our approach, in the supplementary information (SI), we report the modeling results for 24 additional countries that are globally representative, including countries of world

[1]Laboratory for the Modeling of Biological and Socio-technical Systems, Northeastern University, Boston, MA, USA. [2]Networks and Urban Systems Centre, University of Greenwich, London, UK. [3]Department of Epidemiology and Biostatistics, Indiana University School of Public Health, Bloomington, IN, USA. [4]Department of Biostatistics and Bioinformatics, Emory University, Atlanta, USA. [5]ISI Foundation, Turin, Italy. [6]Bruno Kessler Foundation, Trento, Italy. [7]Division of International Epidemiology and Population Studies, Fogarty International Center, National Institutes of Health, Bethesda, MD, USA. [8]Department of Biostatistics, College of Public Health and Health Professions, University of Florida, Gainesville, FL, USA. [9]Fred Hutchinson Cancer Research Center, Seattle, WA, USA. [10]Department of Biostatistics, University of Washington, Seattle, WA, USA. [11]These authors contributed equally: Jessica T. Davis, Matteo Chinazzi, Nicola Perra. ✉e-mail: a.vespignani@northeastern.edu

Article

regions such as Latin America, the Middle East, Africa, East Asia, and Oceania.

In Fig. 1b we show the model estimates of the median daily incidence of new infections up to February 21, 2020, for both the US and Europe. These values are much larger than the number of officially reported cases (see Fig.1a), highlighting the substantial number of potential transmission events that may have already occurred before many states and countries had implemented testing strategies independent of travel history. As validation we compare our model's estimates of the number of infections during the week of March 8, 2020, to the number of cases reported during that week within each US state and European country with at least 1 reported case (shown in Fig.1b inset). While we see a strong correlation between the reported cases and our model's estimated number of infections (Pearson's correlation coefficient on log-values, US: 0.79, $p < 0.001$; Europe: 0.80, $p < 0.001$), far fewer cases had actually been reported by that time. If we assume that the number of reported cases and simulated infections are related through a simple binomial sampling process, we find that on average 9 in 1,000 infections (90%CI [1–35 per 1,000]) and 35 in 1,000 infections (90%CI [4–90 per 1,000]) were detected by March 8, 2020, in the US and Europe. As testing capacity increased, the ascertainment rate grows and our estimates increase to detecting 17 in 1,000 infections (90%CI [2–55 per 1,000]) by March 14, 2020, in the US and 77 in 1,000 infections (90%CI [5–166 per 1,000]) in Europe. The estimated ascertainment rates are in agreement with independent results based on different statistical methodologies[14–16]. In Fig.1c we show the probability that a city in the US or Europe had generated at least 100 infections by February 21, 2020. We see that the progression of the virus through the US and Europe is both temporally and spatially heterogeneous. While many cities had not yet experienced much community transmission by late February, a few areas such as New York City or London very likely already had local outbreaks.

Onset of local transmission

The model's ensemble of realizations provides a statistical description of all the potential pandemic histories compatible with the initial evolution of the pandemic in China. Rather than describing a specific, causal chain of events, we can estimate possible time windows pertaining to the initial of chains of transmission in different geographical regions. We define the onset of local transmission for a country or state as the earliest date when at least 10 new infections are generated per day. This number is chosen because at this threshold the likelihood of stochastic extinction is extremely small[17,18]. As detailed in the SI, further calibration on the US states and European countries suggests posterior values of R_0 ranging from 2.4–2.8. These values are consistent with many other (country dependent) estimates[19–24]. At the same time, given the doubling time of COVID-19 before the implementation of public health measures, any variation of a factor 2 around the 10 infections/day threshold corresponds to a small adjustment of 3–5 days to the presented timelines.

In Fig. 2, we show the posterior probability distribution, $p(t)$, of the week, t, of the onset of local transmission for 15 US states (a) and European countries (b) (see SI for all states/countries). We also calculate, for each country/state, the median date, T, that identifies the first week where the cumulative distribution function is larger than 50%. Among the US states, California and New York state are the earliest, with a date, T, by the week of January 26 (California) and February 2 (New York), 2020. In Europe, Italy, UK, Germany, and France are the first countries with T close to the end of January 2020. However, it is worth noting that each distribution, $p(t)$, has a support spanning several weeks. In Italy, the 5th and 95th percentiles of the $p(t)$ distribution are the week of January 6 and the week of January 30, 2020. These dates also suggest that it is not possible to rule out introductions and transmission events as early as December 2019, although the probability is very small.

For each state in the US and each country in Europe we compared the order in which they surpassed 100 cumulative infections in the model and in the surveillance data (gathered from the John Hopkins University Coronavirus Resource Center[25]). In Extended Data Fig. 1a we plot the ordering for states and compute the Kendall rank correlation coefficient τ (see SI for details). The correlation is positive (τ_{EU}=0.71, $p < 0.001$ and τ_{US}=0.68, $p < 0.001$) indicating that, despite the detection and testing issues, the expected patterns of epidemic diffusion are largely described by the model in both regions.

SARS-CoV-2 introductions

As the model allows the recording of the origin and destination of travelers carrying SARS-CoV-2 at the global scale, we can study the possible sources of SARS-CoV-2 introductions for each US state and European country. More specifically, we record the cumulative number of introductions in each stochastic realization of the model until April 30, 2020. In Fig. 3 we visualize the origin of the introductions considering some key geographical regions (e.g., Europe and Asia) while keeping the US and China separate and aggregating all the other countries (i.e., Others). For both the US and Europe the contribution from mainland China is barely visible and the local share (i.e., sources within Europe and US) becomes significantly higher across the board. Hence, while introduction events in the early phases of the outbreak were key to start the local spreading (see details in the SI), the cryptic transmission phase has been sustained largely by internal flows. Domestic SARS-CoV-2 introductions through April 30, 2020, account for 69% [IQR 60%–81%] of the introductions in California, 78% [IQR 71%–87%] in Texas, and 69% [IQR 60%–80%] in Massachusetts, which is supported by phylogenetic analysis[26]. European origins account for 69% [IQR 60%–80%], 84% [IQR 79%–91%], and 58% [IQR 48%–68%] of the introductions in Italy, Spain, and the UK. In the SI we report the full breakdown for all states and countries.

It is also necessary to distinguish between the full volume of SARS-CoV-2 introductions and the intro- duction events that could be relevant to the early onset of local transmission in each stochastic realization of the model. To this point, it is worth stressing that seeding introductions are different from the actual number of times the virus has been introduced to each location with subsequent onward transmission. Even after a local outbreak has started, future importation events may give rise to additional onward transmission forming independently introduced transmission lineages of the virus[27]. In the model we can investigate seeding events by recording introduction events before the local transmission chains were established. We report the results of this analysis in the SI, showing that importations from mainland China may be relevant in seeding the epidemic in January, but then play a comparatively small role in the COVID-19 expansion in the US and Europe due to the travel restrictions imposed to/from mainland China after January 23, 2020.

The early timing of the initial introductions and early diffusion pattern of SARS-CoV-2 was driven by air travel. We find a positive correlation (τ_{EU}=0.66, $p < 0.001$ and τ_{US}=0.66, $p < 0.001$) comparing the ordering of states according to when they surpassed 100 cumulative, reported cases (referred to as the epidemic order) and their domestic and international air travel volume rank (Extended Data Fig. 1B). Similar observations have been reported in China, where the initial spreading of the virus outside Hubei was strongly correlated with the traffic to/from the province[28]. Other factors like population size are also correlated with both the travel flows (τ_{EU}=0.59, $p < 0.001$ and τ_{US}=0.7, $p < 0.001$) and the epidemic order (τ_{EU}=0.46, $p < 0.001$ and τ_{US}=0.68, $p < 0.001$), which are discussed in detail in the SI. In our model, it is not possible to exclude increased contacts in highly populated places before social distancing interventions and disentangle this effect from increased seeding due to the correlation between travel volume and population size.

COVID-19 burden

Starting in March 2020, the establishment and timing of NPIs as well as other epidemiological drivers (i.e., population size and density, age-structure etc.) determined the disease burden in the US and Europe[29–32]. We account for these features by calibrating the model results, individually, for each US state and European country. More precisely, we estimate the posterior distribution of the infection fatality ratio (IFR) and infection attack rate in each US state and European country. To this end, we adopt the ABC approach using as evidence the number of new deaths reported from March 22, 2020, to June 27, 2020. We consider a uniform prior for the average IFR in the range from 0.4% to 2% that is age stratified proportional to the IFR values reported in Ref. [33]. We also consider a uniform prior for reporting delays between the date of death and reporting ranging from 2 to 22 days in both Europe and the US[34]. Details provided in the SI.

In Fig. 4(A-D, F-I), we report the model fit of the estimated weekly deaths of the first wave for selected states and countries. Additional model results for all investigated regions including a sensitivity analysis of different calibration methods can be found in the SI. We find a strong correlation between the weekly model-estimated deaths and the reported values with a Pearson correlation coefficient of 0.99 ($p < 0.001$) for both Europe and the US (see Fig. S6). As the data suggest, many European countries and US states saw peaks in April and May with various decreasing trajectories that depend on the mitigation strategies in place. Additionally, we report the estimated posteriors for the cumulative infection attack rates and IFRs as of July 4, 2020, in European countries experiencing more than 100 total deaths and the top 20 states ranked by infection attack rate in the US.

Within Europe, Belgium has the highest estimated infection attack rate of 13.2% (90% CI [8.5%-28.3%]) by July 4, 2020, in agreement with the results in Ref. [14]. Furthermore, by that time Belgium reported the highest COVID-19 mortality rate out of the European countries investigated with 8.5 deaths per 10,000 individuals. However, Italy is estimated to have the highest median IFR of 1.4% (90% CI [0.6%-1.8%]), which aligns with other ranges reported in the literature[35,36]. The US states with the highest infection attack rates are located within the Northeast and experienced a significant first wave during March-April 2020. New York and New Jersey are the top two states with infection attack rates of 13.4% (90% CI [9.1% – 26.7%]) and 15.2% (90% CI [10.2% – 31.3%]), respectively. These numbers are aligned with estimates from New York City reported in Ref. [37]. In the SI we report summary tables with estimated IFRs, infection attack rates, as well as the reproductive number in the absence of mitigation measures for all calibrated US states and European countries. Additionally, we compare our attack rate estimates to the prevalence of individuals with SARS-CoV-2 antibodies from serological studies across the US and Europe (Extended Data Fig. 1D). The seroprevalence estimates are compared to the model estimates during the same time window the studies were performed (details on the seroprevalence data from this figure can be found in the Table S8 and Section 9.3 in the SI).

Discussion

The model presented here captures the spatial and temporal heterogeneity of the early stage of the pandemic, going beyond the single country-level reconstruction. It provides a mechanistic understanding of the underlying dynamics of the pandemic's interconnected evolution. Furthermore, rather than showing specific evidence for early infection in a few locations, our study aims at providing a statistical characterization and quantification of the initial transmission pathways at a global scale. Our results can be compared to and complement analyses based on gene sequencing and travel volumes. We find that 72% of the early introductions to Italy, before the local outbreak, are linked to China, which is in agreement with Ref. [38] highlighting the key

role of importations between these regions at the beginning of the pandemic. Additionally, similar to our findings, Ref. [27] estimates that the majority of importation events through April 2020, associated with onward transmission in the UK, came from Europe. The contributions from China are quantified below 1% and limited to the very early phase. Furthermore, seeding events from the US are estimated to be 3% which aligns with our estimate (8% [IQR 3%−9%]). However, their results point to a larger share from Europe (90%) compared to ours (58% [IQR 48%−68%]) and conversely, we estimate a larger contribution from Asia (27% [IQR 19%−35%]). Since our analysis is a statistical description of the possible introduction pathways, differences could arise both due to our model design, but also from genomic sampling biases[39].

The sources of introduction of SARS-CoV-2 infections in Europe and the US changed substantially and rapidly through time. This caused reactive response strategies, such as issuing travel restrictions targeting countries only after local transmission is confirmed, ineffective at preventing local outbreaks. Our results suggest that many regions in the US and Europe experienced an onset of local transmission in January and February 2020, during the time when testing capacity was limited. If testing had been more widespread and not restricted to individuals with a travel history from China, there would have been more opportunities for earlier detection and interventions. In the SI we report a counterfactual scenario where we assume broader testing specifications not based on the individual travel history and find the epidemic progression is considerably delayed (see Section 8 in the SI).

As testing capacity increased and more cases were detected, many governments began to issue social distancing guidelines to mitigate the spread of SARS-CoV-2. The first European country to implement a *cordon sanitaire* was Italy on February 23, 2020, for a few northern cities[40]. Many other countries followed suit and implemented national lockdowns in March, 2020[30,41], however this was weeks after our model estimates that SARS-CoV-2 was introduced and locally spreading. We find a strong correlation between the number of cases reported by the date of a lockdown/social distancing measure and the cumulative infections projected by July, 4, 2020 (Extended Data Fig. 1C), indicating that the earlier NPIs had been issued, the smaller the COVID-19 burden experienced during the first wave. This is in agreement with other analyses showing that the timing of NPIs is crucial in limiting the burden of COVID-19[19,29,42–48]. Overall, our results strengthen the case for preparedness plans with broader indication for testing that are able to detect local transmission earlier.

As with all modeling analyses, results are subject to biases from the limitations and assumptions within the model as well as the data used in its calibration. The model's parameters, such as generation time, incubation period, and the proportion of asymptomatic infections are chosen according to the current knowledge of SARS-CoV-2. Although the model is robust to variations in these parameters (see the SI for the sensitivity analysis), more information on the key characteristics of the disease would considerably reduce uncertainties. The model calibration does not consider correlations among importations (i.e., family travel) and assumes that travel probabilities are age-specific across all individuals in the catchment area of each transportation hub.

In light of the assumptions and limitations inherent to this modeling approach, the results are able to complement the analyses from sequencing data of SARS-CoV-2 genomes used to reconstruct the early epidemic history of the COVID-19 pandemic[38]. The methods used in this analysis offer a blueprint to identify the most likely early spreading dynamics of emerging viruses and they can be used as a real-time risk assessment tool. Anticipating the locations where a virus is most likely to spread to next could be instrumental in guiding enhanced testing and surveillance activities. The estimated SARS-CoV- importation patterns and the cryptic transmission phase dynamics are of potential use when planning and developing public health policies in relation to international traveling and they could provide important insights in assessing the potential risk and impact of emerging SARS-CoV-2

Article

variants in regions of the world with limited testing and genomic surveillance resources.

Online content

Any methods, additional references, Nature Research reporting summaries, source data, extended data, supplementary information, acknowledgements, peer review information; details of author contributions and competing interests; and statements of data and code availability are available at https://doi.org/10.1038/s41586-021-04130-w.

1. CDC. "First Travel-related Case of 2019 Novel Coronavirus Detected in United States". https://www.cdc.gov/media/releases/2020/p0121-novel-coronavirus-travel-case.html (2020).
2. Spiteri, G. et al. First cases of coronavirus disease 2019 (COVID-19) in the WHO European Region, 24 January to 21 February 2020. *Euro surveill*. **25**, 2000178 (2020).
3. Althoff, K. N. et al. Antibodies to SARS-CoV-2 in All of Us Research Program Participants, January 2-March 18, 2020. *Clin. Infect. Dis*. ciab519 (2021). https://doi.org/10.1093/cid/ciab519.
4. La Rosa, G. et al. SARS-CoV-2 has been circulating in northern Italy since December 2019: Evidence from environmental monitoring. *Sci. Total Environ*. **750**, 141711 (2021). https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7428442/
5. Deslandes, A. et al. SARS-CoV-2 was already spreading in France in late December 2019. *Int. J. Antimicrob. Agents* **55**, 106006 (2020). https://www.sciencedirect.com/science/article/pii/S0924857920301643
6. Bedford, T. et al. Cryptic transmission of SARS-CoV-2 in Washington state. *Science* **370**, 571–575 (2020). https://www.sciencemag.org/lookup/doi/10.1126/science.abc0523.
7. Basavaraju, S. V. et al. Serologic Testing of US Blood Donations to Identify Severe Acute Respiratory Syn- drome Coronavirus 2 (SARS-CoV-2)–Reactive Antibodies: December 2019–January 2020. *Clin. Infect. Dis*. **72**, e1004-e1009 (2020). https://doi.org/10.1093/cid/ciaa1785.
8. COVID, C. et al. Evidence for Limited Early Spread of COVID-19 Within the United States, January–February 2020. *MMWR. Morb. Mort. Wkly. Rep*. **69**, 680-684 (2020). https://www.cdc.gov/mmwr/volumes/69/wr/mm6922e1.htm
9. Lescure, F.-X. et al. Clinical and virological data of the first cases of COVID-19 in Europe: a case series. *Lancet Infect. Dis*. **20**, 697-706 (2020). https://doi.org/10.1016/S1473-3099(20)30200-0.
10. Gostic, K., Gomez, A. C., Mummah, R. O., Kucharski, A. J. & Lloyd-Smith, J. O. Estimated effectiveness of symptom and risk screening to prevent the spread of COVID-19. *eLife* **9**, e55570 (2020). https://doi.org/10.7554/eLife.55570.
11. Balcan, D. et al. Multiscale mobility networks and the spatial spreading of infectious diseases. *Proc. Natl. Acad. Sci. U.S.A.* **106**, 21484–21489 (2009). https://doi.org/10.1073/pnas.0906910106.
12. Balcan, D. et al. Modeling the spatial spread of infectious diseases: The GLobal Epidemic and Mobility computational model. *J. Comput. Sci*. **1**, 132–145 (2010). https://doi.org/10.1016/j.jocs.2010.07.002.
13. Sunn°aker, M. et al. Approximate Bayesian Computation. *PLoS Comput. Biol*. **9**, 1–10 (2013). https://doi.org/10.1371/journal.pcbi.1002803.
14. Russell, T. W. et al. Reconstructing the early global dynamics of under-ascertained COVID-19 cases and infections. *BMC Med*. **18**, 332 (2020). https://doi.org/10.1186/s12916-020-01790-9.
15. Gatto, M. et al. Spread and dynamics of the COVID-19 epidemic in Italy: Effects of emergency containment measures. *Proc. Natl. Acad. Sci*. **117**, 10484–10491 (2020). https://doi.org/10.1073/pnas.2004978117.
16. Havers, F. P. et al. Seroprevalence of Antibodies to SARS-CoV-2 in 10 Sites in the United States, March 23-May 12, 2020. *JAMA Intern. Med*. **180**, 1576–1586 (2020). https://doi.org/10.1001/jamainternmed.2020.4130.
17. Bailey, N. T. et al. *The mathematical theory of infectious diseases and its applications* (Charles Griffin & Company Ltd, 5a Crendon Street, High Wycombe, Bucks HP13 6LE., 1975).
18. Colizza, V. & Vespignani, A. Epidemic modeling in metapopulation systems with heterogeneous coupling pattern: Theory and simulations. *J. Theor. Biol*. **251**, 450–467 (2008). https://doi.org/10.1016/j.jtbi.2007.11.028.
19. Perra, N. Non-pharmaceutical interventions during the COVID-19 pandemic: A review. *Phys. Rep*. (2021). https://doi.org/10.1016/j.physrep.2021.02.001.
20. Salje, H. et al. Estimating the burden of SARS-CoV-2 in France. *Science* **369**, 208–211 (2020). https://doi.org/10.1126%2Fscience.abc3517.
21. Domenico, L. D., Pullano, G., Sabbatini, C. E., Bo¨elle, P.-Y. & Colizza, V. Impact of lockdown on COVID-19 epidemic in ˆIle-de-France and possible exit strategies. *BMC Med*. **18** (2020). https://doi.org/10.1186%2Fs12916-020-01698-4.
22. Chang, S. et al. Mobility network models of COVID-19 explain inequities and inform reopening. *Nature* (2020). https://doi.org/10.1038%2Fs41586-020-2923-3.
23. Pei, S., Kandula, S. & Shaman, J. Differential Effects of Intervention Timing on COVID-19 Spread in the United States. *Sci. Adv*. **6**, eabd6370 (2020).
24. Lau, M. S. Y. et al. Characterizing superspreading events and age-specific infectiousness of SARS-CoV-2 transmission in Georgia, USA. *Proc Natl Acad Sci USA* **117**, 22430–22435 (2020). https://doi.org/10.1073%2Fpnas.2011802117.
25. Johns Hopkins University Coronavirus Resource Center. https://coronavirus.jhu.edu/.
26. Fauver, J. R. et al. Coast-to-Coast Spread of SARS-CoV-2 during the Early Epidemic in the United States. *Cell* **181**, 990–996.e5 (2020). https://doi.org/10.1016/j.cell.2020.04.021.
27. du Plessis, L. et al. Establishment and lineage dynamics of the SARS-CoV-2 epidemic in the UK. *Science*. **371**, 708-712 (2021).
28. Kraemer, M. U. et al. The effect of human mobility and control measures on the COVID-19 epidemic in China. *Science* **368**, 493–497 (2020).
29. White, E. R. & H´ebert-Dufresne, L. State-level variation of initial COVID-19 dynamics in the United States. *PLoS ONE* **15**, e0240648 (2020).
30. Desvars-Larrive, A. et al. A structured open dataset of government interventions in response to COVID-19. *Sci Data*. **7**, 285. (2020). https://doi.org/10.1038%2Fs41597-020-00609-9.
31. Althouse, B. M. et al. The unintended consequences of inconsistent pandemic control policies. Preprint at https://www.medrxiv.org/content/10.1101/2020.08.21.20179473v2 (2020).
32. Rader, B. et al. Crowding and the shape of COVID-19 epidemics. *Nat. Med*. **26**, 1829–1834 (2020). https://doi.org/10.1038/s41591-020-1104-0.
33. Verity, R. et al. Estimates of the severity of coronavirus disease 2019: a model-based analysis. *Lancet Infect. Dis*. **20**, 669-677 (2020). https://doi.org/10.1016/S1473-3099(20)30243-7.
34. COVID-19 Pandemic Planning Scenarios, https://www.cdc.gov/coronavirus/2019-ncov/hcp/planning-scenarios.html
35. O'Driscoll, M. et al. Age-specific mortality and immunity patterns of SARS-CoV-2. *Nature*. **590**, 140-145 (2020). https://doi.org/10.1038/s41586-020-2918-0.
36. Poletti, P. et al. Infection fatality ratio of SARS-CoV-2 in Italy. *Euro. Surveill*. **25**, 2001381. (2020). https://doi.org/10.2807/1560-7917.ES.2020.25.31.2001383.
37. Kissler, S. et al. Reductions in commuting mobility correlate with geographic differences in SARS-CoV-2 prevalence in New York City. *Nat. Commun*, **11**, 4674 (2020).
38. Lemey, P. et al. Accommodating individual travel history and unsampled diversity in Bayesian phylogeographic inference of SARS-CoV-2. *Nat. Commun*. **11**, 51120 (2020).
39. Martin, M. A., Van Insberghe, D. & Koelle, K. Insights from SARS-CoV-2 sequences. *Science*. **371**, 466–467 (2021). https://science.sciencemag.org/content/371/6528/466.
40. Lazzerini, M. & Putoto, G. COVID-19 in Italy: momentous decisions and many uncertainties. *Lancet Glob. Health* **8**, e641–e642 (2020).
41. Cheng, C., Barcel´o, J., Hartnett, A. S., Kubinec, R. & Messerschmidt, L. Covid-19 government response event dataset (coronanet v.1.0). *Nat Hum Behav* **4**, 756–768 (2020). https://doi.org/10.1038%2Fs41562-020-0909-7
42. Yang, W., Shaff, J. & Shaman, J. Effectiveness of Non-pharmaceutical Interventions to Contain COVID- 19: A Case Study of the 2020 Spring Pandemic Wave in New York City. *J. R. Soc. Interface*. **18**, 2020082. https://doi.org/10.1098/rsif.2020.0822
43. Ali, S. T. et al. Serial interval of SARS-CoV-2 was shortened over time by nonpharmaceutical interventions. *Science* **369**, 1106–1109 (2020). https://doi.org/10.1126%2Fscience.abc9004.
44. Pan, A. et al. Association of Public Health Interventions with the Epidemiology of the COVID-19 Outbreak in Wuhan, China. *JAMA* **323**, 1915-1923 (2020). https://doi.org/10.1001%2Fjama.2020.6130.
45. Jefferies, S. et al. COVID-19 in New Zealand and the impact of the national response: a descriptive epidemiological study. *Lancet Public Health* **5**, e612–e623. (2020). https://doi.org/10.1016%2Fs2468-2667%2820%2930225-5
46. Auger, K. A. et al. Association Between Statewide School Closure and COVID-19 Incidence and Mortality in the US. *JAMA* **324**, 859-870. (2020). https://doi.org/10.1001%2Fjama.2020.14348.
47. Islam, N. et al. Physical distancing interventions and incidence of coronavirus disease 2019: natural experiment in 149 countries. *BMJ* (2020). https://doi.org/10.1136%2Fbmj.m2743.
48. Haug, N. et al. Ranking the effectiveness of worldwide covid-19 government interventions. *Nat. Hum. Behav*. **4**, 1303–1312 (2020). https://doi.org/10.1038%2Fs41562-020-01009-0.
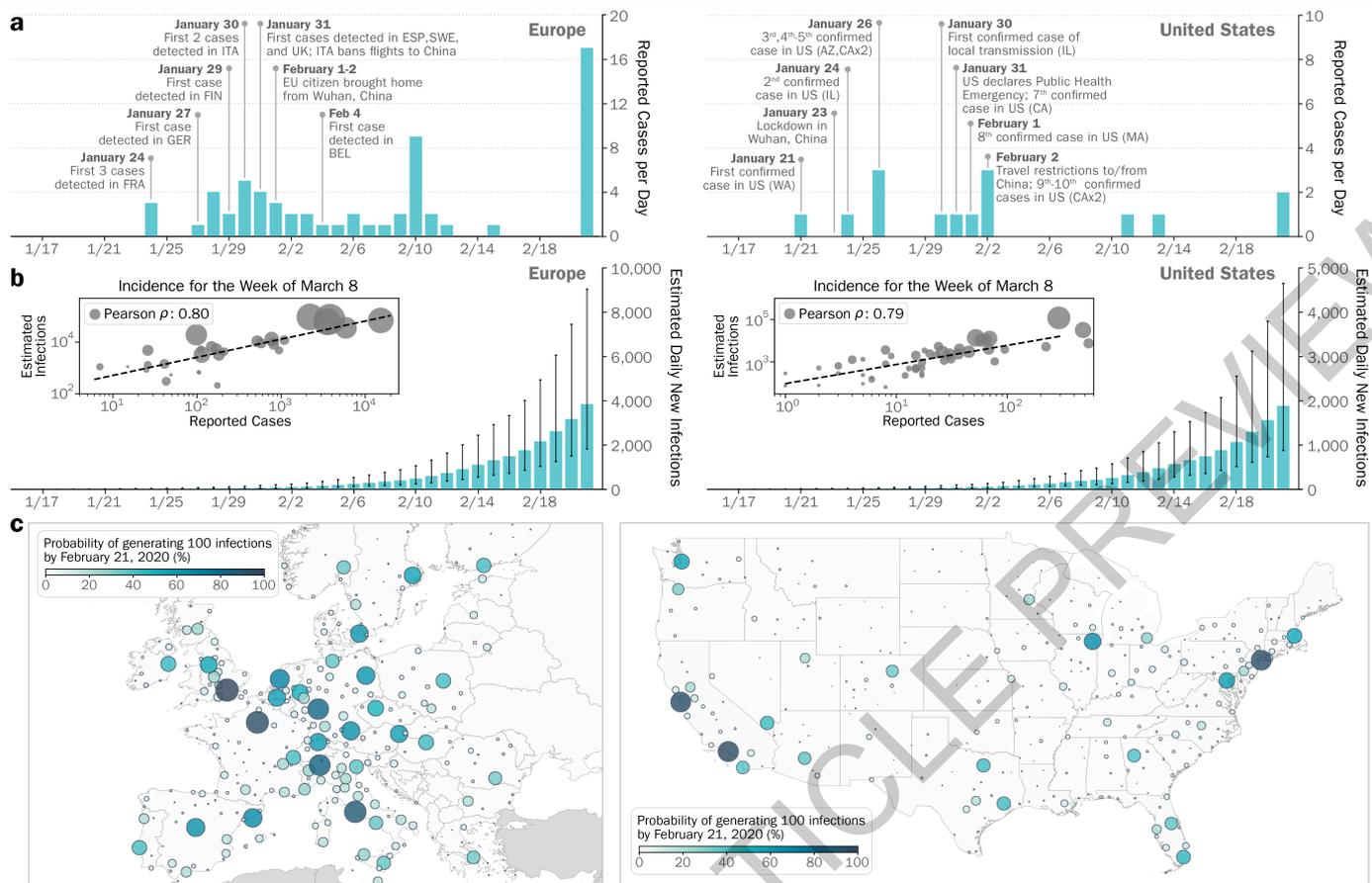
Fig. 1 | Early picture of the COVID-19 outbreak in Europe and the United States. (a) Timelines of the daily reported and confirmed cases of COVID-19 in Europe and US. (b) Model-based estimates for the daily number of new infections in Europe and US. The model estimates reported are the median values with the interquartile range (IQR) obtained with an ABC calibration method using n=200,000 independent model realizations. The inset plot compares the weekly incidence of reported cases with the median, weekly incidence of infections estimated by the model for the week of March 8 – 14, 2020 for the continental US-states and European countries that reported at least 1 case (Europe: n = 30, US: n = 48). Circle size corresponds to the population size of each state and country. (c) The probability that a city in Europe and the US had generated at least 100 cumulative infections by February 21, 2020. Color and circle size are proportional to the probability.
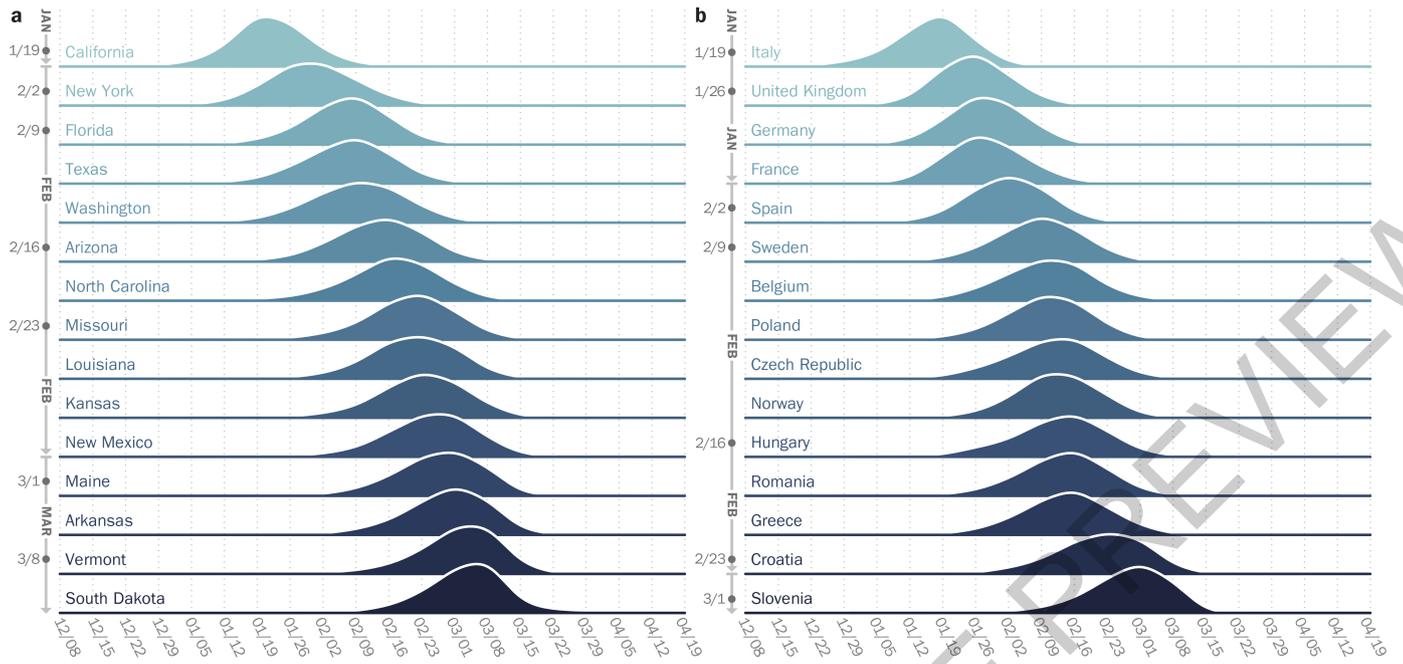
Fig. 2 | Timing of the onset of local transmission. Posterior distributions of the week when each US state (a) or European country (b) first reached 10 locally generated SARS-CoV-2 trans- mission events per day. Countries/states are ordered by the median date of their posterior distribution. The week of this date corresponds to the dates reported on the vertical axis.
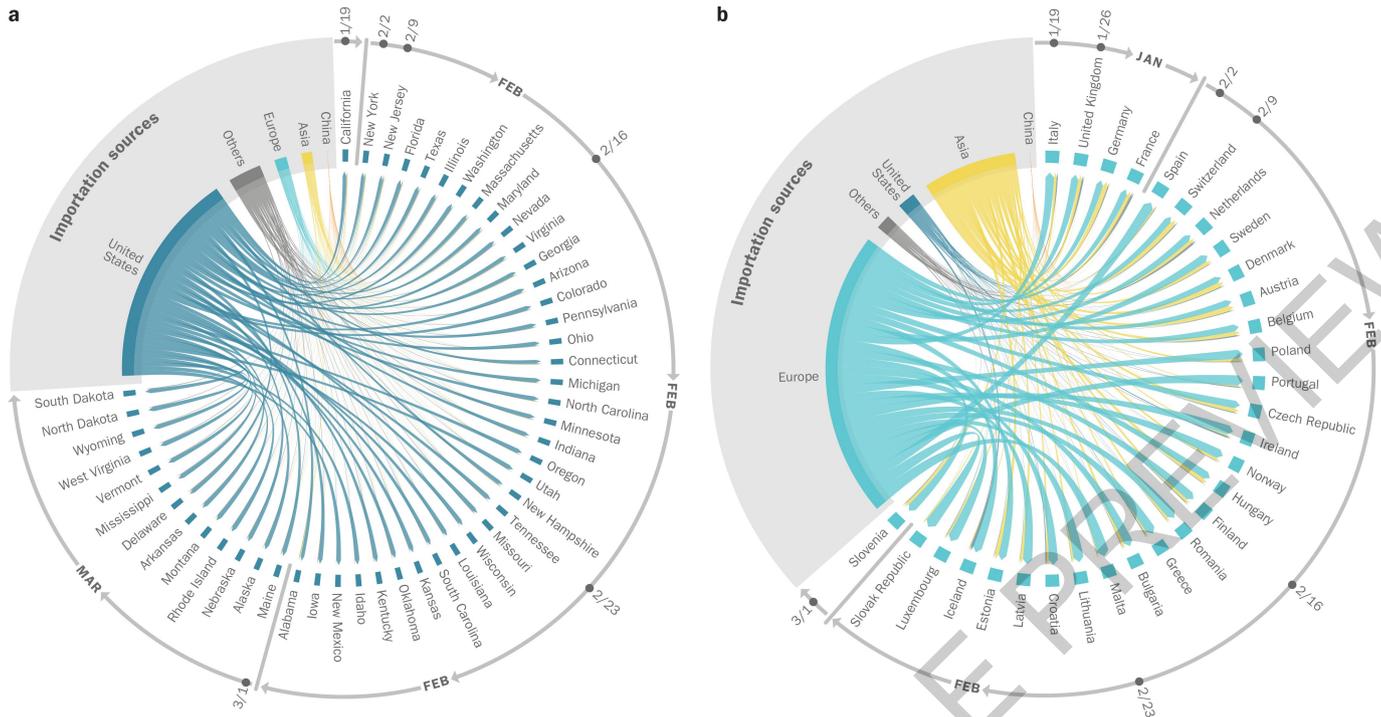
Fig. 3 | Importation sources from the beginning of the outbreak until the end of April 2020. Each US state (a) and European country (b) is displayed in a clockwise order with respect to the start of the local outbreak (as seen in Fig. 2). Importation flows are directed and weighted. We normalize links considering the total in-flow for each state so that the sum of importations flows, for each state, is one. In the SI we report the complete list of countries contributing as importation sources in each geographical region.
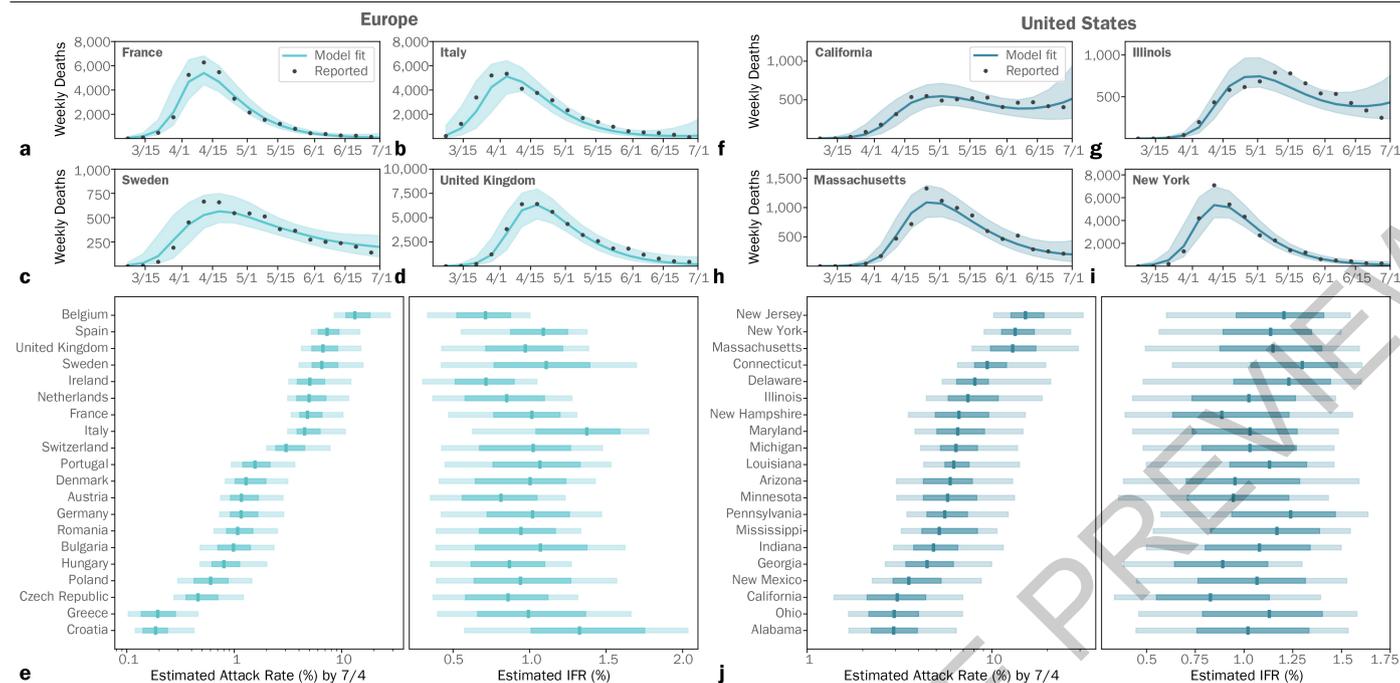
Fig. 4 | The burden of the first wave in Europe and the US. (a-d) Model fit of the estimated weekly deaths for selected countries in Europe. (e) Posterior distributions of the infection attack rates and infection fatality ratios by July 4, 2020, for European countries where there were at least 100 reported deaths. (f-i) Model fit of the estimated weekly deaths for selected states in the US. (j) Posterior distributions of the estimated infection attack rates and infection fatality ratios by July 4, 2020 for the top 20 US states (ranked according to their infection attack rates). The curves in (a-d,f-i) show the median values and the 90%CIs. For e and j the outer, lighter boxes represent the 90%CI, the darker, inner boxes represent the IQR, and the vertical lines represent the median value. Posterior distributions in e and j are the result of the ABC analysis of 200,000 independent model realizations.

Methods

Global Epidemic and Mobility Model

The Global Epidemic and Mobility Model, GLEAM, is a stochastic, spatial, age-structured metapopulation model. Previously this model was used to characterize the early stage of the COVID-19 epidemic in mainland China to estimate the effectiveness of travel bans and restrictions[49]. GLEAM divides the global population into more than 3,200 subpopulations in roughly 200 different countries and territories inter-connected by realistic air-travel and commuting mobility networks. A subpopulation is defined as the catchment area around major transportation hubs. The airline transportation data encompass daily travel data in the origin-destination format from the Official Aviation Guide (OAG) database[50] reflecting actual traffic changes that occurred during the pandemic. Ground mobility and commuting flows are derived from the analysis and modeling of data collected from the statistics offices of 30 countries on five continents[11,12]. The international travel data account for travel restrictions and government issued policies. Furthermore, the model accounts for the reduction of internal, country-wide mobility and changes in contact patterns in each country and state in 2020. Specific model details are reported in the SI.

SARS-CoV-2 transmission dynamic

The transmission dynamics take place within each subpopulation and assume a classic SLIR-like compartmentalization scheme for disease progression similar to those used in several large-scale models of SARS-CoV-2 transmission[15,51–55]. Each individual, at any given point in time, is assigned to a compartment corresponding to their particular disease-related state (being, e.g., susceptible, latent, infectious, removed)[49]. This state also controls the individual's ability to travel (details in the SI). Individuals transition between compartments through stochastic chain binomial processes. Susceptible individuals can acquire the virus through contacts with individuals in the infectious category and can subsequently become latent (i.e., infected but not yet able to transmit the infection). The process of infection is modeled using age-stratified contact patterns at the state and country level[56,57]. Latent individuals progress to the infectious stage at a rate inversely proportional to the latent period, and infectious individuals progress to the removed stage at a rate inversely proportional to the infectious period. The sum of the mean latent and infectious periods defines the generation time. Removed individuals are those who can no longer infect others. To estimate the number of deaths we consider a uniformly distributed prior of the infection fatality ratios (ranging from 0.4% to 2%) that is age stratified proportional to the values estimated by Ref.[33] and incorporates reporting delays. The transmission model does not assume heterogeneities due to age differences in susceptibility to the SARS-CoV-2 infection for younger children (1-10 years old). This is an intense area of discussion[58–60,61–70]. The transmission dynamic and the offspring distribution of infectious individuals in the model will depend on the specific details of each population, local and global mobility, NPIs etc. While overdispersion in transmission varies by location in our model, we find that overall, it is consistent with 25% of primary infections causing 75% of transmission in our simulations (Fig. S9). Additional simulations considering a fixed level of dispersion, informed by past studies, result in differences of less than 3 days in onset times (Fig. S10; see also the SI for further discussion).

Reporting summary

Further information on research design is available in the Nature Research Reporting Summary linked to this paper.

Data availability

Epidemic surveillance data were collected from the Johns Hopkins Coronavirus Resource Center https://coronavirus.jhu.edu/. Proprietary airline data are commercially available from OAG (https://www.oag.com/) and IATA (https://www.iata.org/) databases. Other model intervention data includes Google's COVID-19 Com- munity Mobility Reports available at https://www.google.com/covid19/mobility/ and the Oxford COVID-19 Response Tracker available at https://github.com/OxCGRT/covid-policy-tracker. Source data are provided with this paper.

Code availability

The GLEAM model is publicly available at http://www.gleamviz.org/. All data analyses of model results were performed using python v3.8 Source data are provided with this paper.

49. Chinazzi, M. et al. The effect of travel restrictions on the spread of the 2019 novel coronavirus (COVID-19) outbreak. *Science* **368**, 395–400 (2020).
50. Official Aviation Guide https://www.oag.com/.
51. Natsuko Imai, Anne Cori, Ilaria Dorigatti, Marc Baguelin, Christl A. Donnelly, Steven Riley, Neil M. Ferguson. Report 3: Transmissibility of 2019-nCoV. https://www.imperial.ac.uk/mrc-global-infectious-disease-analysis/covid-19/report-3-transmissibility-of-covid-19/ (2020).
52. Kissler, S. M., Tedijanto, C., Goldstein, E., Grad, Y. H. & Lipsitch, M. Projecting the transmission dynamics of SARS-CoV-2 through the postpandemic period. *Science* **368**, 860–868 (2020). https://science.sciencemag.org/content/368/6493/860.
53. Li, R. et al. Substantial undocumented infection facilitates the rapid dissemination of novel coronavirus (SARS- CoV-2). *Science* **368**, 489–493 (2020). https://science.sciencemag.org/content/368/6490/489.
54. Wu, J. T., Leung, K. & Leung, G. M. Nowcasting and forecasting the potential domestic and international spread of the 2019-nCoV outbreak originating in Wuhan, China: a modelling study. *Lancet*. **395**, 10225 (2020).
55. Lai, S. et al. Effect of non-pharmaceutical interventions to contain COVID-19 in China. *Nature*. **585**, 410-413 (2020). https://doi.org/10.1038/s41586-020-2293-x.
56. Mistry, D. et al. Inferring high-resolution human mixing patterns for disease modeling. *Nature communications* **12**, 1–12 (2021).
57. Prem, K., Cook, A. R. & Jit, M. Projecting social contact matrices in 152 countries using contact surveys and demographic data. *PLoS Comput. Biol.* **13**, e1005697 (2017).
58. Zhang, J. et al. Changes in contact patterns shape the dynamics of the COVID-19 outbreak in China. *Science*. **368**, 1481-1486 (2020). https://science.sciencemag.org/content/early/2020/04/28/science.abb8001.
59. Davies, N. G. et al. Age-dependent effects in the transmission and control of COVID-19 epidemics. *Nat. Med*. **26**, 1205–1211 (2020). https://doi.org/10.1038/s41591-020-0962-9.
60. Bi, Q. et al. Epidemiology and transmission of COVID-19 in 391 cases and 1286 of their close contacts in Shenzhen, China: a retrospective cohort study. *Lancet Infect. Dis.* **20**, 911-919 (2020). https://doi.org/10.1016/S1473-3099(20)30287-5.
61. Rambaut, A. "Preliminary phylogenetic analysis of 11 nCoV2019 genomes, 2020-01-19" (2020). http://virological.org/t/preliminary-phylogenetic-analysis-of-11-ncov2019-genomes-2020-01-19/329
62. Anderson, K., "Clock and TMRCA based on 27 genomes"(2020); http://virological.org/t/clock-and-tmrca-based-on-27-genomes/347
63. Bedford, T. et al., "Genomic analysis of nCoV spread. Situation report 2020-01-23" (2020); https://nextstrain.org/narratives/ncov/sit-rep/2020-01-23.
64. Pekar, J., Worobey, M., Moshiri, N., Scheffler, K. & Wertheim, J. O. Timing the SARS-Cov-2 index case in Hubei province. *Science*. **372**, 412–417 (2021). https://science.sciencemag.org/content/372/6540/412.
65. van Dorp, L. et al. Emergence of genomic diversity and recurrent mutations in SARS-CoV-2. *Infect. Genet. Evol.* **83**, 104351 (2020). https://www.sciencedirect.com/science/article/pii/S1567134820301829
66. De Salazar, P. M., Niehus, R., Taylor, A., Buckee, C. & Lipsitch, M. Identifying Locations with Possible Undetected Imported Severe Acute Respiratory Syndrome Coronavirus 2 Cases by Using Importation Predictions. *Emerg. Infect. Dis.* **26**, 7 (2020).
67. Niehus, R., De Salazar, P. M., Taylor, A. R. & Lipsitch, M. Using observational data to quantify bias of traveller-derived COVID-19 prevalence estimates in Wuhan, China. *Lancet Infect. Dis.* **20**, 803– 808 (2020).
68. Global security index; https://www.ghsindex.org/.
69. Nadeau, S. A., Vaughan, T. G., Scire, J., Huisman, J. S. & Stadler, T. The origin and early spread of SARS- CoV-2 in *Europe. Proc. Natl Acad. of Sci. U.S.A.* **118** (2021). https://www.pnas.org/content/118/9/e2012008118.
70. Worobey, M. et al. The emergence of SARS-CoV-2 in Europe and North America. *Science*. **370**, 6516 (2020). https://science.sciencemag.org/content/early/2020/09/09/science.abc8169.
71. Kontis, V. et al. Magnitude, demographics, and dynamics of the effect of the first wave of the COVID-19 pandemic on all-cause mortality in 21 industrialized countries. *Nat. Med*. **26**, 1919–1928 (2020). https://doi.org/10.1038/s41591-020-1112-0.
72. Trump issues 'Coronavirus Guidelines' for next 15 days to slow pandemic. https://www.cnbc.com/2020/03/16/trumps-coronavirus-guidelines-for-next-15-days-to-slow-pandemic.html.

Article

Extended Data Fig. 1 | Correlation Analysis for European countries and US states. (a) The correlation between the ordering of each country/state to reach 100 infections in the model-estimates and to reach 100 reported cases in the surveillance data (Europe: n = 23, US: n = 49). (b) The correlation between the ordering of each country/state considering the time needed to reach 100 reported cases in the surveillance data and the ranking of the combined international and domestic air traffic (Europe n = 23, US n = 49). Correlations in (a,b) are computed considering the Kendall rank correlation coefficient reported with a two-sided p-value, we consider European countries that reached at least 100 reported deaths by July 4, 2020 and countries in Scandinavia (c) Left: the correlation between the number of cases reported by the date of lockdown for European countries (from Table 4 in Ref.[71]) and the estimated total number of infections by July 4, 2020 (median values, n = 15). Right: the correlation between the number of cases reported by March 16, 2020

(the date the "15 days to slow the spread" guidelines were released in the US Ref.[72]) for each US state and the estimated total infections by July 4, 2020 (median values, n = 36). We consider states that reached at least 100 reported deaths by July 4, 2020. The circle sizes in (a-c) correspond to the population sizes of each country/state. (d) The correlation between the model-estimated infection attack rate and the serological prevalence collected from studies, n = 20. Estimated attack rates are the posterior distributions that are the result of the ABC analysis of 200,000 independent model realizations. Data points refer to different dates and the locations for which serological surveys were available (see table S8 in SI for study descriptions). The model-estimated attack rates use the median value, and the error bars represent the 90%CI. The correlations are calculated using the Pearson correlation coefficient in (c-d) reported with a two-sided p-value.

Article

Extended Data Table 1 | Regions under investigation

United States	Europe
Alabama, Alaska, Arizona, Arkansas, California, Colorado, Connecticut, Delaware, Florida, Georgia, Idaho, Illinois, Indiana, Iowa, Kansas, Kentucky, Louisiana, Maine, Maryland, Massachusetts, Michigan, Minnesota, Mississippi, Missouri, Montana, Nebraska, Nevada, New Hampshire, New Jersey, New Mexico, New York, North Carolina, North Dakota, Ohio, Oklahoma, Oregon, Pennsylvania, Rhode Island, South Carolina, South Dakota, Tennessee, Texas, Utah, Vermont, Virginia, Washington, West Virginia, Wisconsin, Wyoming	Austria, Belgium, Bulgaria, Croatia, Czech Republic, Denmark, Estonia, Finland, France, Germany, Greece, Hungary, Iceland, Ireland, Italy, Latvia, Lithuania, Luxembourg, Malta, Netherlands, Norway, Poland, Portugal, Romania, Slovak Republic, Slovenia, Spain, Sweden, Switzer- land, United Kingdom

List of European countries and US states analyzed.

nature portfolio

Corresponding author(s):   Alessandro Vespignani

Last updated by author(s):   Oct 4, 2021

Reporting Summary

Nature Portfolio wishes to improve the reproducibility of the work that we publish. This form provides structure for consistency and transparency in reporting. For further information on Nature Portfolio policies, see our Editorial Policies and the Editorial Policy Checklist.

Statistics

For all statistical analyses, confirm that the following items are present in the figure legend, table legend, main text, or Methods section.

n/a	Confirmed	
☒	☐	The exact sample size (*n*) for each experimental group/condition, given as a discrete number and unit of measurement
☒	☐	A statement on whether measurements were taken from distinct samples or whether the same sample was measured repeatedly
☐	☒	The statistical test(s) used AND whether they are one- or two-sided *Only common tests should be described solely by name; describe more complex techniques in the Methods section.*
☒	☐	A description of all covariates tested
☒	☐	A description of any assumptions or corrections, such as tests of normality and adjustment for multiple comparisons
☐	☒	A full description of the statistical parameters including central tendency (e.g. means) or other basic estimates (e.g. regression coefficient) AND variation (e.g. standard deviation) or associated estimates of uncertainty (e.g. confidence intervals)
☐	☒	For null hypothesis testing, the test statistic (e.g. *F*, *t*, *r*) with confidence intervals, effect sizes, degrees of freedom and *P* value noted *Give P values as exact values whenever suitable.*
☐	☒	For Bayesian analysis, information on the choice of priors and Markov chain Monte Carlo settings
☒	☐	For hierarchical and complex designs, identification of the appropriate level for tests and full reporting of outcomes
☒	☐	Estimates of effect sizes (e.g. Cohen's *d*, Pearson's *r*), indicating how they were calculated

Our web collection on statistics for biologists contains articles on many of the points above.

Software and code

Policy information about availability of computer code

Data collection	No software was used for data collection
Data analysis	The GLEAM model is publicly available at http://www.gleamviz.org/. All data analyses of model results were performed using python v3.8

For manuscripts utilizing custom algorithms or software that are central to the research but not yet described in published literature, software must be made available to editors and reviewers. We strongly encourage code deposition in a community repository (e.g. GitHub). See the Nature Portfolio guidelines for submitting code & software for further information.

Data

Policy information about availability of data

All manuscripts must include a data availability statement. This statement should provide the following information, where applicable:

- Accession codes, unique identifiers, or web links for publicly available datasets
- A description of any restrictions on data availability
- For clinical datasets or third party data, please ensure that the statement adheres to our policy

Epidemic surveillance data were collected from the Johns Hopkins Coronavirus Resource Center https://coronavirus.jhu.edu/. Proprietary airline data are commercially available from OAG (https://www.oag.com/) and IATA (https://www.iata.org/) databases. Other model intervention data includes Google's COVID-19 Community Mobility Reports available at https://www.google.com/covid19/mobility/ and the Oxford COVID-19 Response Tracker available at https://github.com/OxCGRT/covid-policy-tracker.

Field-specific reporting

Please select the one below that is the best fit for your research. If you are not sure, read the appropriate sections before making your selection.

☒ Life sciences   ☐ Behavioural & social sciences   ☐ Ecological, evolutionary & environmental sciences

For a reference copy of the document with all sections, see nature.com/documents/nr-reporting-summary-flat.pdf

Life sciences study design

All studies must disclose on these points even when the disclosure is negative.

Sample size	We use all available data generated by model simulations. We do not generate primary biological or epidemiological data from field experiments.
Data exclusions	No data were excluded
Replication	All data used are described in the data availability statement. Model generated data were generated synthetically using the GLEAM tool documented here: http://www.gleamviz.org/simulator/GLEAMviz_client_manual_v7.0.pdf
Randomization	N/A. We did not perform/consider individual subject studies.We did not allocate any individuals to control or experimental groups.
Blinding	N/A. We did not perform/consider individual subject studies. We did not allocate any individuals to control or experimental groups.

Reporting for specific materials, systems and methods

We require information from authors about some types of materials, experimental systems and methods used in many studies. Here, indicate whether each material, system or method listed is relevant to your study. If you are not sure if a list item applies to your research, read the appropriate section before selecting a response.

Materials & experimental systems

n/a	Involved in the study
☒ ☐	Antibodies
☒ ☐	Eukaryotic cell lines
☒ ☐	Palaeontology and archaeology
☒ ☐	Animals and other organisms
☒ ☐	Human research participants
☒ ☐	Clinical data
☒ ☐	Dual use research of concern

Methods

n/a	Involved in the study
☒ ☐	ChIP-seq
☒ ☐	Flow cytometry
☒ ☐	MRI-based neuroimaging