Genetic drift versus regional spreading dynamics of COVID-19

R. Di Pietro [1], M. Basile [1], L. Antolini [2] and S. Alberti [3 *]

[1] Department of Medicine and Aging Sciences, Section of Biomorphology, G. d'Annunzio University of Chieti-Pescara, Italy.

[2] Center for Biostatistics, Department of Clinical Medicine, Prevention and Biotechnology, University of Milano-Bicocca, Monza, Italy.

[3] Unit of Medical Genetics, Department of Biomedical Sciences - BIOMORF, University of Messina, via Consolare Valeria, Messina, Italy.

[*] Address reprint requests to Prof. Alberti at the Unit of Medical Genetics, Department of Biomedical Sciences - BIOMORF, University of Messina, via Consolare Valeria, Messina, Italy, or at salberti@unime.it

Drs. Di Pietro and Basile contributed equally to this article.

Running title: COVID-19 diffusion rate determinants

Word count of the abstract: 236

Word count of the text: 2,472

Abstract

Background

Current propagation models of COVID-19 pandemic spreading appear poorly consistent with existing epidemiological data and with evidence that SARS-CoV-2 is rapidly mutating, for potential aggressive evolution of the disease.

Methods

We thus challenged environmental versus genetic evolution models of COVID-19 spreading, in the largest epidemiological analysis conducted as yet. This was performed over 168,089 laboratory-confirmed infection cases in Italy, Spain and Scandinavia. Landmark dates were set for each of the countries analyzed at peak diffusion rates this date, and the doubling time versus cumulative number of diagnoses was computed. Diffusion data in Germany, France and UK constituted the validation dataset of our model, over 210,239 additional cases. Mutations and mutation rates of SARS-CoV-2 versus COVID-19 spreading were analyzed at nextstrain.org/ncov/europe.

Results

The mean doubling time of COVID-19 was 6.63 days in northern Italy, 5.87 days in central areas, and 5.38 days in southern Italy, for shorter COVID-19 doubling time in warmer regions. Spain extended this trend, with a mean COVID-19 doubling time of 4.2 days. Slower diffusion across progressively colder regions was observed in Scandinavia, with 9.4 days COVID-19 doubling time in Sweden, 10.8 days in Finland and 12.95 days in Norway. Such diffusion model was supported by SARS-CoV-2 mutation strings upon sequential diffusion across the North/South gradient.

Conclusions

Our findings indicate COVID-19 association to a sharp North/South climate gradient, with faster spreading in southern regions. Thus, warmer climate conditions may not limit SARS-CoV-2 diffusion. Very cold regions may be better spared by recurrent courses of SARS-CoV-2 infection.

Keywords: COVID-19, pandemic, spreading dynamics, climate areas, mutation rates.

Introduction

A first study on 425 cases identified initial transmission dynamics of Severe Acute Respiratory Syndrome (SARS) in 2019 (COVID-19) in China (1). In its early stages, the epidemic doubled in size every 6.4 (2) to 7.4 (1) days, with a reproductive number (R_0) of infectious cases from 2.2 (1) to 2.7 (2). Later studies described how the disease spread to Singapore (3), then to Germany (4), France and Finland (www.ecdc.europa.eu/en/covid-19-pandemic) (5-7).

However, key epidemiological evidence remained to be acquired (7). Major uncertainties remained on COVID-19 spreading determinants. SARS-CoV-2 was proposed to be sensitive to temperature and humidity, which may affect diffusion across diverse climate areas (8) (papers.ssrn.com/sol3/papers.cfm?abstract_id=3550308; ssrn.com/abstract=3556998; www.medrxiv.org/content/10.1101/2020.02.22.20025791v1). Diversity among substrains of SARS-CoV-2 occurs across different regions in the world (nextstrain.org/ncov/global). SARS-CoV-2 possesses a single-strand RNA genome (9) and was soon found to acquire genomic mutations. Selective pressure may apply to SARS-CoV-2 genomic drifting, and this may cooperatively drive geographic diffusion.

Current propagation models predicted a limited impact of COVID-19 in the Southern hemisphere during seasons that were infection-prone in the Northern hemisphere (papers.ssrn.com/sol3/papers.cfm?abstract_id=3550308; ssrn.com/abstract=3556998). However, early foci of infection were detected in Australia and New Zealand (Figure 1). Outbreaks were also revealed in South America and extended to Central America and Mexico. Further infection foci were revealed in Saudi Arabia and Africa, and extended to sub-Saharan countries (Tables S1, S2), questioning simple models of climate-dependent COVID-19 spreading. Coronaviruses spread to some extent similarly to the influenza virus (8), through small droplets suspended in the air, suggesting sensitivity to environmental humidity and temperature conditions. A recent meta-analysis (10), though, indicated resilience of coronaviruses to the environment. In a comparison to SARS-CoV-1, SARS-CoV-2 remained viable in aerosols for hours, and persisted over solid surfaces, 72 hours on plastic, 48 hours on stainless steel and 24 hours on cardboard (11), raising issues on current SARS-CoV-2 diffusion models.

This led us to challenge a genetic versus climate-driven additive coronavirus infection model. A robust analysis of SARS-CoV-2 spreading determinants required high-information density (12, 13). Case incidence models depend on complex factors interplay (global traveling, founder effect versus time from initial infection (2), population clustering in big cities, social dynamics, infectious ability of the virus (14, 15), COVID-19 containment procedures). Among them, a major confounding factor is the time of initial infection at any given place, which, everything else being equal, leads to

vastly different absolute numbers of derived cases (2). Velocity of infection spreading had previously been shown to be a hard composite index of the R_0 of the virus and of patients viral load/disease stage/severity (1, 16) and is insensitive to the time of infection seeding. Hence, infection doubling time was utilized as a COVID-19 descriptor in this study.

Vastly diverse climatic regions around the CET longitude (15ºE), were severely exposed to infection. Spain and Italy were the countries with the highest initial incidence of COVID-19 in Europe (Figures 1, S1, Table S3). The heaviest initial casualties in Italy were suffered by Lombardy and Veneto, i.e. cold and humid areas during wintertime. Markedly warmer and drier climate conditions prevail in southern regions of the country. A further shift toward warmer/drier conditions occurs in Spain. Scandinavian countries appeared initially spared by the infection (Table S4) and provided a reference for cold winter temperatures, over a Sweden-Finland-Norway axis. Thus, we assessed a climate-dependent coronavirus infection model, through a global-scale analysis of 86,498 infection cases in Italy, 64,095 in Spain, versus 17,496 cases in Scandinavia (github.com/pcm-dpc/COVID-19) (Supplemental Appendix). Diffusion data in France (Table S5), Germany (Table S6), and UK (Table S7) were utilized as a validation dataset of 210,239 infection cases. This model was then merged with the coronavirus genetic drift-driven diffusion determinants, according to mutation trajectories in the analyzed areas.

Methods

Incidence data

Laboratory-confirmed infection cases in Europe cases were retrieved at peak diffusion rates as follows: Italy (github.com/pcm-dpc/COVID-19, March 27th 2020), France (dashboard.covid19.data.gouv.fr/vue-d-ensemble?locatio n=FRA; April 4th 2020), UK (www.nhs.uk/; April 9th 2020), Germany (corona.rki.de; April 2nd 2020), Spain (RTVE - Ministry of Health; www.rtve.es/noticias/20200415/mapa-del-coronavirus-espana/2004681.shtml; March 31st 2020), Sweden (Public Health Agency of Sweden; www.folkhalsomyndigheten.se/smittskydd-beredskap/utbrott/aktuella-utbrott/covid-19; April 13th 2020), Finland (National Institute for Health and Welfare THL; thl.fi/en/web/thlfi-en; April 7th 2020), Norway; data from the Norwegian Institute of Public Health; www.fhi.no/sv/smittsomme-sykdommer/corona/dags--og-ukerapporter/dags--og-ukerapporter-om-koronavirus; April 7th 2020). All available data in each national registry were systematically included in the analysis.

Incidence data were collapsed into a global database, to explore case incidence over time, and health outcome measures across countries and country provinces. Disease severity was classified as (a) hospitalized cases, (b) intensive-care unit patients, (c) recovered cases, (d) deaths. Incidence

scatter plots by region were linked to Köppen–Geiger climate classification maps (koeppen-geiger.vu-wien.ac.at/present.htm). These were computed as mean parametrization of 1980-2016 data (17). The three-variable classification by country areas was quantified as a string and utilized as an independent variable versus COVID-19 spreading velocity (Table 1).

SARS-CoV-2 mutation analysis

SARS-CoV-2 genomic RNA sequences and country-correlated data were obtained from nextstrain.org/ncov/global. Scatter plots were generated, by strings of acquired mutations over time and overall number of mutations per genome per chosen area. Phylogeny trees for compiled mutations strings were then obtained according to mutant branch descriptors (nextstrain.org/ncov/europe?branchLabel=aa) (Figures S2-9).

Statistical analysis

The cumulative incidence of COVID-19 diagnoses was contrasted to calendar time for each province in a scatter plot (12, 13). These plots acted as a smoother, for determining the trajectory of infection cases. A landmark date for total numbers of diagnoses was set according to case incidence shape in each dataset. From this date, the doubling time for cumulative number of diagnoses was calculated backward for each province as follows. Two dates were identified: the maximum date, at which the cumulative number of diagnoses were lower than a half of the cumulative number of diagnoses at the landmark time, and the minimum date, with a cumulative number of diagnoses greater than a half of the cumulative number of diagnoses at the landmark date. The fraction of days from the minimum date to achieve half of the cumulative number of diagnoses at the landmark date were obtained by a linear assumption for the cumulative incidence between the two dates. Comparison of doubling time values was conducted versus central intercepts. Coefficients, standard error, 95% confidence intervals were computed. Percentile distribution boxplots of COVID-19 doubling times were drawn. Median, maximum value, minimum value and distribution outliers were computed. The correlation of discrete values of COVID-19 spreading rates curve versus climate-area string values was computed by Anova.

Software

Stata software version 16 was used for data importing, manipulation and graphics (StataCorp. 2019. *Stata Statistical Software: Release 16*. College Station, TX: StataCorp LLC).

Results

Our attention was first drawn to the Southern hemisphere. Simple propagation models predicted essential absence of COVID-19 diffusion, during seasons that were infection-prone in the Northern hemisphere. However, early foci of infection were detected in Australia and New Zealand (Figure 1*A*). South and Central America appeared initially spared. Assessment at later time points indicated, though, large-scale (≥30 infection cases) outbreaks in Argentina, Bolivia, Brazil, Chile, Colombia, Ecuador, Perù, Uruguay, Venezuela. Parallel outbreaks were revealed in Costa Rica, Dominican Republic, Panama and Mexico (Figure 1*C*, Table S1).

Africa, Middle-East and the Arabian peninsula also appeared spared during the initial course of COVID-19 (Figure 1*A*). However, infection foci appeared soon in Saudi Arabia, a non-high-risk country by most standards. This was soon recognized as a risk for COVID-19 spreading (18), and Saudi Arabia suspended the Umrah pilgrimage to Mecca and Medina on March 4th. Additional cases were reported in United Arab Emirates, Bahrain, Kuwait, Oman. Infectious foci were revealed in other countries facing the Persian Gulf and the Gulf of Oman, such as Iraq, Iran, Afghanistan, Pakistan. Further outbreaks were recorded in continental Africa, i.e. in Algeria, Egypt, Burkina Faso, Senegal, Democratic Republic of the Congo, Cameroon, Côte d'Ivoire, Ghana, Nigeria, South Africa (Table S2).

COVID-19 doubling time by geographic area

We thus moved on to computing spreading rates of COVID-19 in compared areas in Europe. Global data were collected from country registries and infection rates over time were computed for:

Italy: on infection cases from March 3rd to March 27th (n=86,498) (Supplemental Appendix) (Figures S10-12).

Spain: on infection cases from February 25th to March 27th 2020 (n=64,095) (Figure S13).

Norway: data (>50 infection case outbreaks) were obtained from February 21st to April 14th 2020 (n=6,676) (Figure S14).

Finland: on infection cases from March 1st to April 7th 2020 (n=2,646) (Figure S15).

Sweden: data (>50 infection case outbreaks) were obtained from February 26th to April 9th 2020 (n=8,995) (Figure S16).

France: on infection cases from February 25th to April 4th 2020 (Figure S17).

UK: on infection cases from February 1st to April 9th 2020 (Figure S17).

Germany: on infection cases from February 24th to April 2nd 2020 (Figure S17).

COVID-19 doubling times by Countries, Regions and Provinces were computed as indicated. Landmark dates were utilized as set for each analyzed geographic area. From this date the time for doubling the cumulative number of diagnoses was calculated backward for each province.

COVID-19 doubling time versus climate region

Quantitative climate assessments are affected by interdependent sets of variables, such as humidity and temperature, which provide sources of uncertainty in climate models (19). We thus utilized the Köppen–Geiger climate classification maps (koeppen-geiger.vu-wien.ac.at/present.htm), as drawn over 30+ years of observations. This was distilled as a three-variable classification by country areas, quantified as a string and utilized as an independent variable versus COVID-19 spreading velocity (Table 1).

Summary doubling times were grouped by geographic region. The average doubling time for northern Italy was 6.63 (SD=1.94) days; 5.87 (SD=1.08) days in central regions; 5.38 (SD=2.31) days in southern areas, for significantly shorter doubling time in southern regions (P=0.02) (Table S3, Figures S10-12). The mean COVID-19 doubling-time for the whole country was 6.06 (SD=1.95) days (Table S3).

With a doubling time of 4.2-days, Spain extended such an indication (Figure S13). At the opposite end of the climate spectrum, Scandinavia showed longer COVID-19 doubling times, over a Sweden-Finland-Norway axis, with a doubling time of 9.4 days (SD=1.2) for Sweden (P<0.0001 versus northern Italy), 10.8 days for Finland, 12.95 days (SD=0.52) for Norway (P<0.0001 versus northern Italy) (Table S3, Figures S14-16). This depicted a distinct North-South gradient of COVID-19 spreading velocity (Anova P<0.0001) (Table 1).

Such climate model was challenged versus COVID-19 diffusion rates in Germany, France and UK over 210,239 laboratory-confirmed infection cases. Pandemic doubling time was computed to be 7.0 days in Germany (Figure S17). In sharp consistency, those in France and UK were 7.5 and 7.2 days, respectively. Average climate areas for Germany, France and UK were Cfb Köppen–Geiger climate classification classes (Table 1), which bridged classification classes of Northern Italy and Southern Sweden, as predicted by the model.

SARS-CoV-2 genetic-drift driven diffusion

The SARS-CoV-2 genomic RNA was shown to progressively mutate over time (nextstrain.org/ncov/europe). To determine whether mutation strings correlated with diffusion at distinct geographic areas, SARS-CoV-2 genomic RNA sequences from different countries were obtained at sequential times. Scatter plots were then generated, for strings of acquired mutations versus overall number of accumulated mutations versus time. Phylogeny trees for compiled mutations were then obtained, according to mutant clade descriptors (nextstrain.org/ncov/europe?branchLabel=aa).

Sequence mutation analysis revealed up to seven major branches of linear mutation acquisition, at sites of major diffusion after spreading from China (Figures S2-9). A mutation string-driven aggressiveness of SARS-CoV-2 spreading, was predicted to lead to (a) correlation of specific strings with hardest-hit countries, (b) a late predominance of one/few dominant strings over the course of COVID-19 and (c) increase in disease severity over time. The highest numbers of accumulated mutations were revealed in SARS-CoV-2 in Wales and Senegal isolates, which were identified as late disease correlates. Consistent, the lowest number of accumulated mutations was recorded in Italy, the country that was first to show strong disease severity in Europe. A large mutation load was observed in Spain (n=14), the second hardest-hit country in Europe, as close to that of Sweden (n=13), a country with much more limited COVID-19 diffusion. Large mutation loads in late-disease-insurgence countries, such as France and Belgium (n=16), lend additional support to drive by duration of disease course. Consistent, four major mutation strings branches were revealed in all examined European countries, indicating dependence on specific substrains of SARS-CoV-2.

Discussion

Rapid COVID-19 diffusion in Southern hemisphere countries, such as Australia, New Zealand, South and Central America, together with early infection outbreaks in Africa, Middle-East and in the Arabian peninsula posed early questions on influenza-like propagation models of SARS-CoV-2. Potent models were subsequently constructed that better took the complexity of COVID-19 diffusion into account (19, 20). However, actual, high-dimensional data on COVID-19 infection dynamics remained largely missing (7).

Our work provided the needed data to this end, as the largest epidemiological analysis of COVID-19 diffusion conducted as yet. All available data from each national registry were systematically included in the analysis, for over 378,328 laboratory-confirmed infection cases in continental Europe and UK. This analysis was then merged with mutation-string-driven SARS-CoV-2 spreading at distinct geographic areas.

Four major mutation strings branches were revealed in all examined European countries, indicating relationship with specific substrains of SARS-CoV-2. The highest numbers of accumulated mutations were revealed in SARS-CoV-2 in Wales and Senegal isolates, which were identified as late disease correlates. The lowest number of accumulated mutations was recorded in Italy, the country that was first to show severe disease outbreaks in Europe. A large mutation load was observed in Spain, which followed as second hardest-hit country in Europe, as close to that of Sweden, a country with late COVID-19 diffusion. The largest mutational loads were revealed in France and Belgium, as late-disease-insurgence countries, further supporting intertwining with successive stages

of disease course.

We then found that COVID-19 spreading velocity followed a North-South gradient in Italy, for significantly shorter doubling times in southern regions. With a doubling time of 4.2-days, Spain extended such an indication. At the opposite end of the climate spectrum, Scandinavia showed longer COVID-19 doubling times, over a Sweden-Finland-Norway axis, for a sharp, quantitative North-South gradient of COVID-19 spreading velocity. This climate model was verified in validation datasets of COVID-19 diffusion in Germany, France and UK over 210,239 laboratory-confirmed infection cases. Pandemic doubling times were sharply consistent in Germany, France and UK, according to ultimate climate-area Köppen–Geiger fingerprints, thus coordinately bridging Northern Italy classes with Southern Sweden.

These results indicate significantly faster diffusion of COVID-19 in warmer regions. Such findings are consistent with indications of resilience of coronaviruses to climate conditions (10) and long-term viability of SARS-CoV-2 in the environment (11). Of note, the Middle East Respiratory Syndrome (MERS) was first reported in Saudi Arabia (www.cdc.gov/coronavirus/mers). MERS is caused by the MERS-CoV, which is structurally and genetically related to SARS-CoV. MERS is endemic in the Arabic Peninsula, indicating that at least specific coronavirus strains are resilient to desert climate conditions (www.cdc.gov/coronavirus/mers/risk.html).

Taken together, our findings suggest resilience of SARS-CoV-2 in warmer regions, and caution that high environmental temperatures may not efficiently tame SARS-CoV-2 infectiousness (20). On the other hand, very cold regions may be better spared by recurrent courses of COVID-19.

Acknowledgments

We are much indebted to all the information curators we cite, and to the website providers the article data and graphic primers have been downloaded from.

Role in the article

All authors contributed to literature search, figures, study design, data collection, data analysis, data interpretation. S.A and R.DiP. wrote the manuscript draft. All authors contributed to discussing and writing the final text. R.DiP. and M.B. contributed equally to this work.

Footnote page

Conflict of interest

The authors do not have a commercial or other association that might pose a conflict of interest.

Funding

The University of Messina and Oncoxx Biotech srl provided funding to this work.

Meeting presentation

These findings have not yet been presented to meetings.

Corresponding author contact information

Prof. Saverio Alberti, Head, Medical Genetics, Department of Biomedical Sciences, University of Messina, via Consolare Valeria, 98100 Messina, Italy, Phone: (+39) 090-221.3375, E-mail: salberti@unime.it

299 **References**

300 1. Li Q, Guan X, Wu P, Wang X, Zhou L, Tong Y, Ren R, Leung KSM, Lau EHY, Wong JY, Xing
301 X, Xiang N, Wu Y, Li C, Chen Q, Li D, Liu T, Zhao J, Liu M, Tu W, Chen C, Jin L, Yang R, Wang
302 Q, Zhou S, Wang R, Liu H, Luo Y, Liu Y, Shao G, Li H, Tao Z, Yang Y, Deng Z, Liu B, Ma Z,
303 Zhang Y, Shi G, Lam TTY, Wu JT, Gao GF, Cowling BJ, Yang B, Leung GM, Feng Z. 2020.
304 Early Transmission Dynamics in Wuhan, China, of Novel Coronavirus–Infected
305 Pneumonia. New England Journal of Medicine 382:1199-1207.
306 2. Wu JT, Leung K, Leung GM. 2020. Nowcasting and forecasting the potential domestic and
307 international spread of the 2019-nCoV outbreak originating in Wuhan, China: a
308 modelling study. The Lancet 395:689-697.
309 3. Pung R, Chiew CJ, Young BE, Chin S, Chen MIC, Clapham HE, Cook AR, Maurer-Stroh S,
310 Toh MPHS, Poh C, Low M, Lum J, Koh VTJ, Mak TM, Cui L, Lin RVTP, Heng D, Leo Y-S, Lye
311 DC, Lee VJM, Kam K-q, Kalimuddin S, Tan SY, Loh J, Thoon KC, Vasoo S, Khong WX,
312 Suhaimi N-A, Chan SJH, Zhang E, Oh O, Ty A, Tow C, Chua YX, Chaw WL, Ng Y, Abdul-
313 Rahman F, Sahib S, Zhao Z, Tang C, Low C, Goh EH, Lim G, Hou Ya, Roshan I, Tan J, Foo K,
314 Nandar K, Kurupatham L, Chan PP, et al. 2020. Investigation of three clusters of COVID-
315 19 in Singapore: implications for surveillance and response measures. The Lancet
316 395:1039-1046.
317 4. Hoehl S, Rabenau H, Berger A, Kortenbusch M, Cinatl J, Bojkova D, Behrens P,
318 Böddinghaus B, Götsch U, Naujoks F, Neumann P, Schork J, Tiarks-Jungk P, Walczok A,
319 Eickmann M, Vehreschild MJGT, Kann G, Wolf T, Gottschalk R, Ciesek S. 2020. Evidence
320 of SARS-CoV-2 Infection in Returning Travelers from Wuhan, China. New England
321 Journal of Medicine 382:1278-1280.
322 5. Sun J, He W-T, Wang L, Lai A, Ji X, Zhai X, Li G, Suchard MA, Tian J, Zhou J, Veit M, Su S.
323 2020. COVID-19: Epidemiology, Evolution, and Cross-Disciplinary Perspectives. Trends
324 in Molecular Medicine doi:https://doi.org/10.1016/j.molmed.2020.02.008.
325 6. McMichael TM, Currie DW, Clark S, Pogosjans S, Kay M, Schwartz NG, Lewis J, Baer A,
326 Kawakami V, Lukoff MD, Ferro J, Brostrom-Smith C, Rea TD, Sayre MR, Riedo FX, Russell
327 D, Hiatt B, Montgomery P, Rao AK, Chow EJ, Tobolowsky F, Hughes MJ, Bardossy AC,
328 Oakley LP, Jacobs JR, Stone ND, Reddy SC, Jernigan JA, Honein MA, Clark TA, Duchin JS.
329 2020. Epidemiology of Covid-19 in a Long-Term Care Facility in King County,
330 Washington. New England Journal of Medicine doi:10.1056/NEJMoa2005412.
331 7. Lipsitch M, Swerdlow DL, Finelli L. 2020. Defining the Epidemiology of Covid-19 —
332 Studies Needed. New England Journal of Medicine 382:1194-1196.

8. Sundell N, Andersson L-M, Brittain-Long R, Lindh M, Westin J. 2016. A four year seasonal survey of the relationship between outdoor climate and epidemiology of viral respiratory tract infections in a temperate climate. Journal of Clinical Virology 84:59-63.
9. Mousavizadeh L, Ghasemi S. 2020. Genotype and phenotype of COVID-19: Their roles in pathogenesis. Journal of Microbiology, Immunology and Infection doi:https://doi.org/10.1016/j.jmii.2020.03.022.
10. Kampf G, Todt D, Pfaender S, Steinmann E. 2020. Persistence of coronaviruses on inanimate surfaces and their inactivation with biocidal agents. Journal of Hospital Infection 104:246-251.
11. van Doremalen N, Bushmaker T, Morris DH, Holbrook MG, Gamble A, Williamson BN, Tamin A, Harcourt JL, Thornburg NJ, Gerber SI, Lloyd-Smith JO, de Wit E, Munster VJ. 2020. Aerosol and Surface Stability of SARS-CoV-2 as Compared with SARS-CoV-1. New England Journal of Medicine doi:10.1056/NEJMc2004973.
12. Ambrogi F, Biganzoli E, Querzoli P, Ferretti S, Boracchi P, Alberti S, Marubini E, Nenci I. 2006. Molecular subtyping of breast cancer from traditional tumor marker profiles using parallel clustering methods. Clin Cancer Res 12:781-90.
13. Cimoli G, Malacarne D, Ponassi R, Valenti M, Alberti S, Parodi S. 2004. Meta-analysis of the role of p53 status in isogenic systems tested for sensitivity to cytotoxic antineoplastic drugs. Biochim Biophys Acta 1705:103-20.
14. To KK-W, Tsang OT-Y, Leung W-S, Tam AR, Wu T-C, Lung DC, Yip CC-Y, Cai J-P, Chan JM-C, Chik TS-H, Lau DP-L, Choi CY-C, Chen L-L, Chan W-M, Chan K-H, Ip JD, Ng AC-K, Poon RW-S, Luo C-T, Cheng VC-C, Chan JF-W, Hung IF-N, Chen Z, Chen H, Yuen K-Y. 2020. Temporal profiles of viral load in posterior oropharyngeal saliva samples and serum antibody responses during infection by SARS-CoV-2: an observational cohort study. The Lancet Infectious Diseases doi:https://doi.org/10.1016/S1473-3099(20)30196-1.
15. Chen Y, Li L. 2020. SARS-CoV-2: virus dynamics and host response. The Lancet Infectious Diseases doi:10.1016/S1473-3099(20)30235-8.
16. Liu Y, Eggo RM, Kucharski AJ. 2020. Secondary attack rate and superspreading events for SARS-CoV-2. The Lancet 395:e47.
17. Beck HE, Zimmermann NE, McVicar TR, Vergopolan N, Berg A, Wood EF. 2018. Present and future Köppen-Geiger climate classification maps at 1-km resolution. Scientific data 5:180214-180214.
18. Ebrahim SH, Memish ZA. 2020. COVID-19: preparing for superspreader potential among Umrah pilgrims to Saudi Arabia. The Lancet 395:e48.

19. Baker RE, Yang W, Vecchi GA, Metcalf CJE, Grenfell BT. 2020. Susceptible supply limits the role of climate in the early SARS-CoV-2 pandemic. Science doi:10.1126/science.abc2535:eabc2535.

20. Kissler SM, Tedijanto C, Goldstein E, Grad YH, Lipsitch M. 2020. Projecting the transmission dynamics of SARS-CoV-2 through the postpandemic period. Science doi:10.1126/science.abb5793:eabb5793.

374 **Figure legends**

375

376 **Figure 1. Worldwide progression of COVID-19.**

377 (***A***) COVID-19 case incidence, as of 21st 2020; numbers are indicated by color code and circle
378 diameter (www.healthmap.org/covid-19/).

379 (***B***) COVID-19 case incidence, as in (***A***), zoomed over Central Europe.

380 (***C***) COVID-19 incidence of active cases, as of March 31st, 2020; numbers are indicated by circle
381 diameter (Johns Hopkins University, JHU; coronavirus.jhu.edu/map.html).

382 (***D***) Coronavirus spreading around the world as of April 4th. Confirmed cases by country since
383 February 24th (JHU, public.flourish.studio/visualisation/1694807/).

384

385 **Figure 2. COVID-19 diffusion and SARS-CoV-2 mutations.**

386 (*A*) SARS-CoV-2 virus spread according to mutation load.

387 (*B*) Radial diagram of SARS-CoV-2 mutations worldwide. Concentric circles correspond to the
388 number of genomic mutations detected in individual virus isolates.

389

390 **Figure 3. COVID-19 diffusion across geographic areas.**

391 (*top*) Distribution boxplots of COVID-19 doubling times. Upper horizontal line: 75th percentile; lower
392 horizontal line: 25th percentile; horizontal bar within box: median; upper horizontal bar outside box:
393 maximum value; lower horizontal bar outside box: minimum value. Dots: distribution outliers.
394 (*bottom*) doubling time values versus central intercept – Northern Italy. Coef.: coefficient; Std. Err.:
395 standard error; 95% confidence intervals are shown. P>t: 0.002 Southern versus Northern Italy;
396 <0.0001 Sweden versus Northern Italy; <0.0001 Norway versus Northern Italy.

397

398 **Figure 4. The COVID-19 North-South gradient.**

399 The COVID-19 North-South doubling-time gradient across countries by climate zone is depicted.
400 The Anova P-value for trend of the curve is shown. Vertical arrows: COVID-19 doubling times in
401 validation datasets (Germany, France, UK).

402

403 **Table 1: COVID-19 doubling time versus climate area.**

Country/region	COVID-19 doubling time (days)	Climate area	Lab-confirmed case numbers *
Spain	4.2	Csa/Csb/Bsk	64,095
Southern Italy	5.38	Csa/Csb	5,322
Central Italy	5.87	Csa/Cfa/Cfb	10,842
Northern Italy	6.63	Cfa/Cfb	70,334
Germany	7.0	Cfb	73,522
France	7.5	Cfb	68,665
UK	7.2	Cfb	68,052
Sweden	9.4	Dfc/Cfb	11,321
Finland	10.8	Dfc/Dfb	2,646
Norway	12.95	Dfc/Dfb/ET	5,855

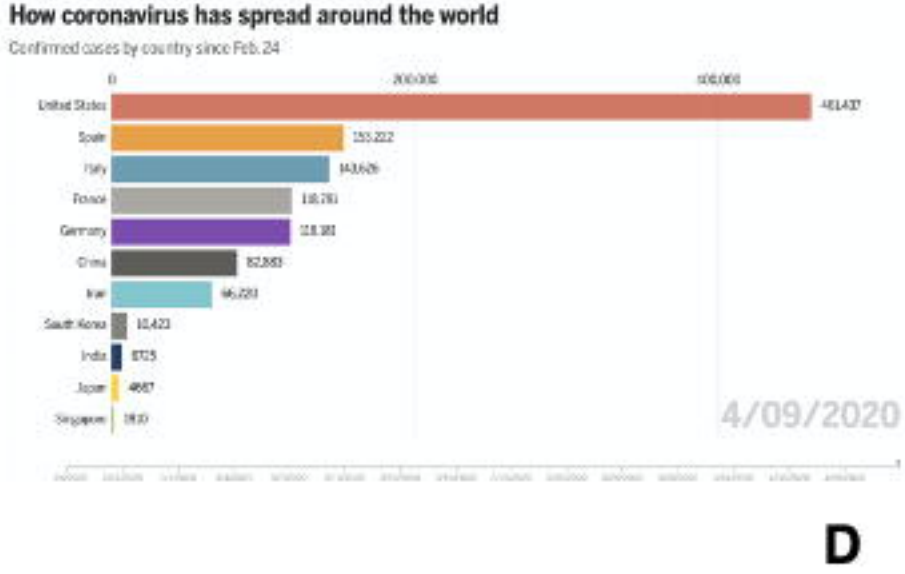
404

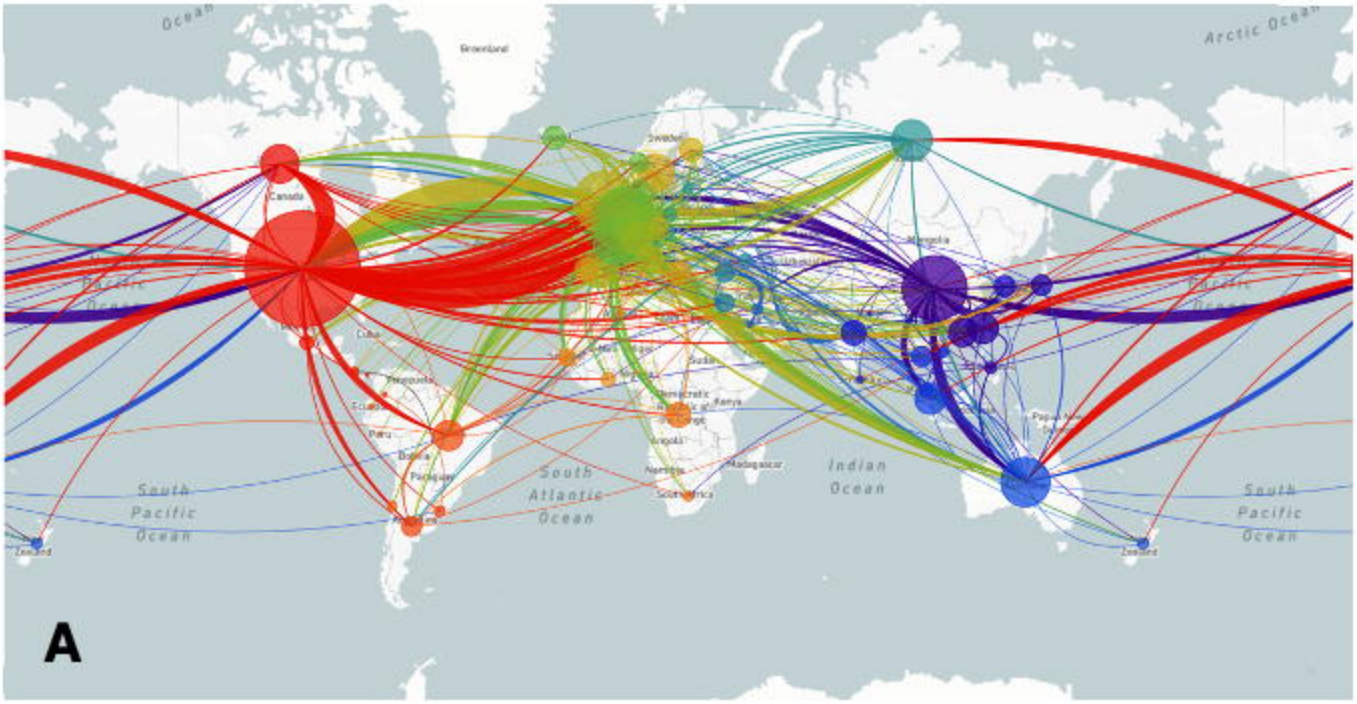
405 *: Laboratory-confirmed infection cases in Europe cases were retrieved by country at peak diffusion rates as follows:
406 Italy (github.com/pcm-dpc/COVID-19, March 27th 2020), France (dashboard.covid19.data.gouv.fr/vue-d-
407 ensemble?locatio n=FRA; April 4th 2020), UK (www.nhs.uk/; April 9th 2020), Germany (corona.rki.de; April 2nd 2020),
408 Spain (RTVE - Ministry of Health; www.rtve.es/noticias/20200415/mapa-del-coronavirus-espana/2004681.shtml; March
409 31st 2020), Sweden (Public Health Agency of Sweden; www.folkhalsomyndigheten.se/smittskydd-
410 beredskap/utbrott/aktuella-utbrott/covid-19; April 13th 2020), Finland (National Institute for Health and Welfare THL;
411 thl.fi/en/web/thlfi-en; April 7th 2020), Norway; data from the Norwegian Institute of Public Health;
412 www.fhi.no/sv/smittsomme-sykdommer/corona/dags--og-ukerapporter/dags--og-ukerapporter-om-koronavirus).

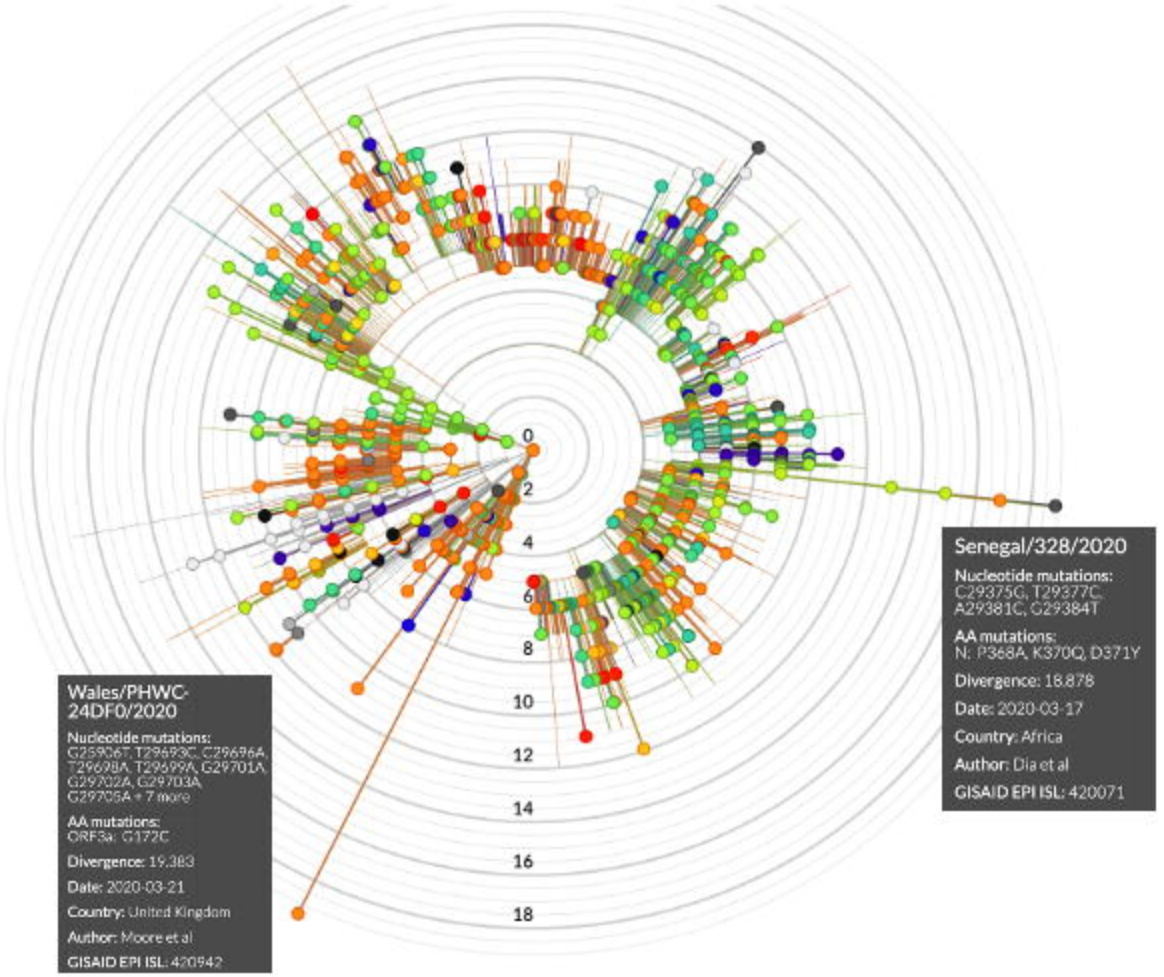

3/21/2020
Number of cases: 50+ 25-49 10-24 < 10
A

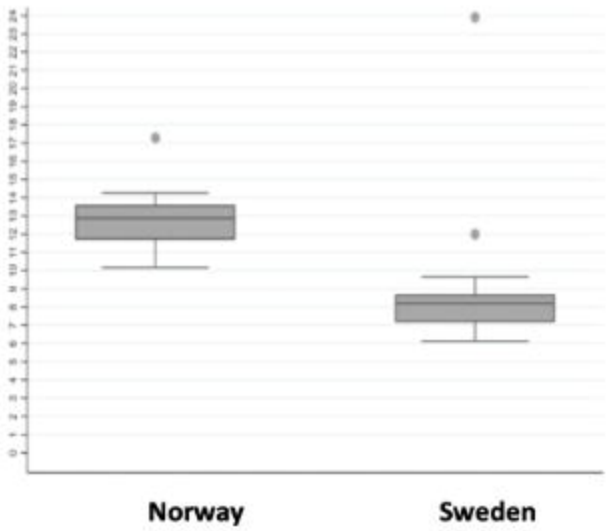

3/31/2020
EUROPE
ASIA
AFRICA
SOUTH AMERICA
AUSTRALIA
C

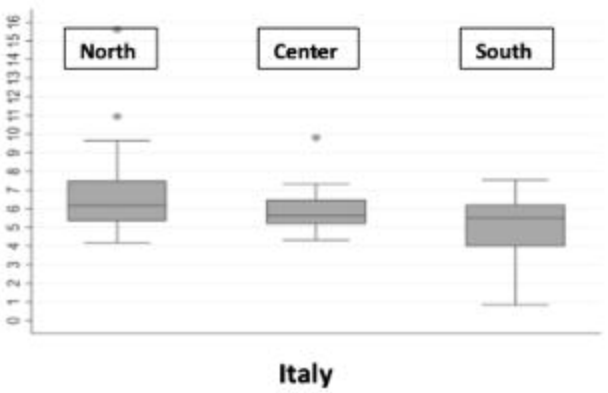

3/21/2020
Number of cases: 50+ 25-49 10-24 < 10
B


How coronavirus has spread around the world
Confirmed cases by country since Feb. 24
0
200,000
400,000
United States
Spain
Italy
France
Germany
China
Iran
South Korea
India
Japan
Singapore
4/09/2020
D

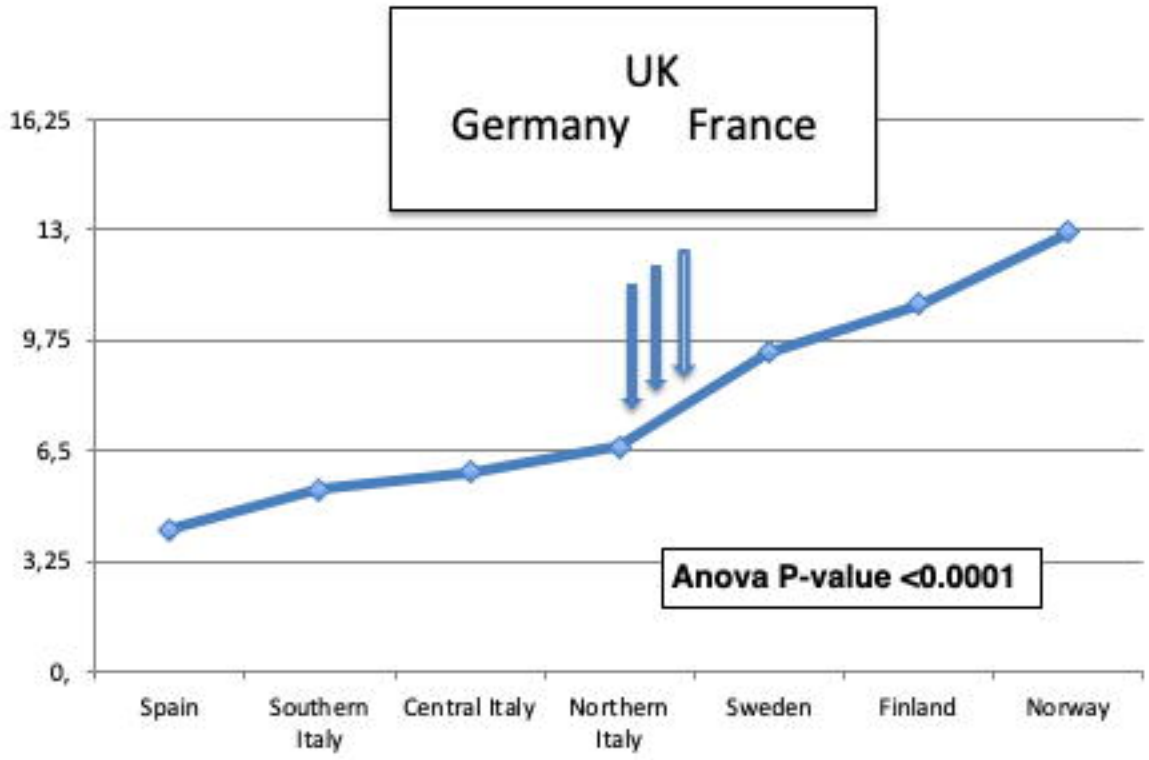

Arctic Ocean
Greenland
Canada
Mongolia
South Pacific Ocean
South Atlantic Ocean
Indian Ocean
Madagascar
Namibia
South Africa
Kenya
Peru
Bolivia
Paraguay
Chile
Ecuador
Papua New Guinea
New Zealand
A


0
2
4
6
8
10
12
14
16
18
Wales/PHWC-24DF0/2020
Nucleotide mutations: G259067, T29693C, C29696A, T29698A, T29699A, G29701A, G29702A, G29703A, G29705A + 7 more
AA mutations: ORF3a: G172C
Divergence: 19.383
Date: 2020-03-21
Country: United Kingdom
Author: Moore et al
GISAID EPI ISL: 420942
Senegal/328/2020
Nucleotide mutations: C29375G, T29377C, A29381C, G29384T
AA mutations: N: P368A, K370Q, D371Y
Divergence: 18.878
Date: 2020-03-17
Country: Africa
Author: Dia et al
GISAID EPI ISL: 420071
B





Region	Coef.	Std. Err.	t	P>t	[95% Conf.	Interval]
intercept Northern Italy	6,63	0,33			5,97	7,29
Southern vs Northern Italy	−1,25	0,53	−2,35	**0,02**	−2,31	−0,20
Central vs Northern Italy	−0,76	0,55	−1,39	**0,166**	−1,84	0,32
Sweden vs Northern Italy	2,77	0,72	3,86	**<0.0001**	1,35	4,19
Norway vs Northern Italy	6,32	0,77	8,25	**<0.0001**	4,80	7,83


UK
Germany France
16,25
13,
9,75
6,5
3,25
0,
Anova P-value <0.0001
Spain
Southern Italy
Central Italy
Northern Italy
Sweden
Finland
Norway