bioRxiv preprint doi: https://doi.org/10.1101/2021.01.12.426373; this version posted January 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

## Title:

# Phylogenetic analyses of SARS-CoV-2 B.1.1.7 lineage suggest a single origin followed by multiple exportation events versus convergent evolution

## Authors

Vrancken B.[1],Dellicour S.[2,3], Smith D.M.[4], Chaillon A[4]

## Affiliations

[1]Laboratory of Clinical and Evolutionary Virology, Department of Microbiology, Immunology and Transplantation, Rega Institute, Katholieke Universiteit Leuven, Leuven, Belgium

[2]Spatial Epidemiology Lab (SpELL), Université Libre de Bruxelles, CP160/12, 50 av. FD Roosevelt, 1050 Bruxelles, Belgium

[3]Department of Microbiology, Immunology and Transplantation, Rega Institute, Laboratory for Clinical and Epidemiological Virology, KU Leuven - University of Leuven, Leuven, Belgium

[4] Division of Infectious Diseases and Global Public Health, University of California San Diego, CA, USA.

WITHDRAWN

see manuscript DOI for details

## Corresponding Authors

Antoine Chaillon

achaillon@health.ucsd.edu

Davey M Smith

d13smith@health.ucsd.edu

Abstract. The emergence of new variants of SARS-CoV-2 herald a new phase of the pandemic. This study used state-of-the-art phylodynamic methods to ascertain that the rapid rise of B.1.1.7 "Variant of Concern" most likely occurred by global dispersal rather than convergent evolution from multiple sources.

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.12.426373; this version posted January 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

WITHDRAWN
see manuscript DOI for details

Following phylogenetic and epidemiological investigations, the SARS-CoV-2 genetic lineage B.1.1.7 is suspected to be associated with an increase human-to-human viral transmissibility[1,2], and was classified as a “Variant of Concern” (VOC B.1.1.7) on December 18, 2020[3]. The variant was first discovered in Kent, United Kingdom (UK) on September 21, 2020, and has since been identified in over 29 countries across the world, including the United States[3,4]. We sought to evaluate whether the breadth of VOC B.1.1.7 identification represents convergent evolution[5] or rapid local and global dispersal after this lineage's genesis.

On January 7’ 2021, we downloaded all B.1.1.7 lineage SARS-CoV-2 genomic sequences available on the GISAID public database[6] (8,786 full length genome sequences across 29 countries, **Supplementary Table 1**). The vast majority were from the UK (96.4%, *n*=8468), but 318 sequences were from other countries, including 13 from North America (8 from USA and 5 from Canada; **Figure 1**).

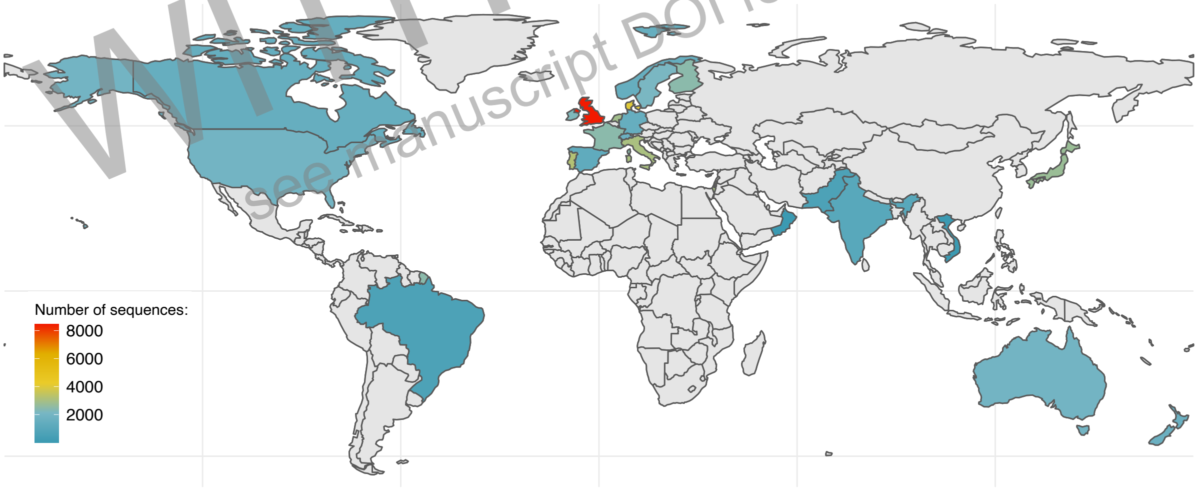
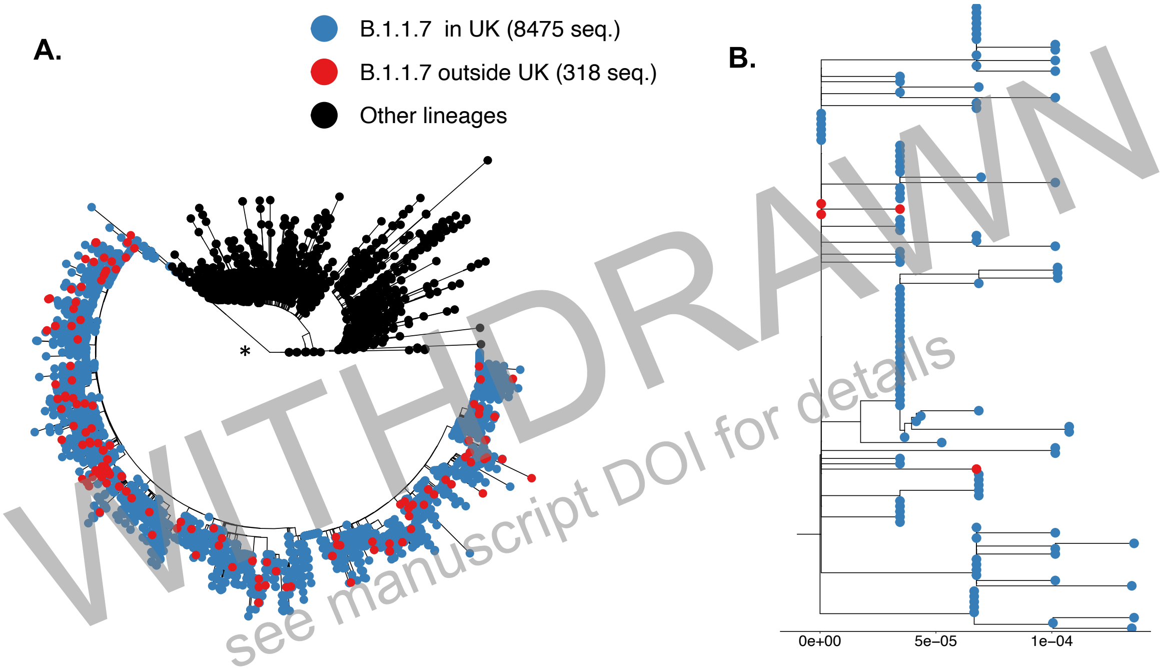


**Figure 1. Map of the B.1.1.7 genomic sequences available on GISAID as of January 7, 2020.** Countries are colored based on the number of publicly available B.1.1.7 sequences.

We combined these B.1.1.7 sequences with a representative set of non-B.1.1.7 sequences (*n*=3,163) based on sequence homology (see Supplementary). The final set of 11,949 sequences was aligned with MAFFT[7] and a Maximum Likelihood phylogeny was inferred using IQ-TREE v2.1.2[8]. The resulting phylogeny showed that all available B.1.1.7 samples cluster together with high support (0.99 Shimodaira Hasegawa [SH] support[9-11]). Non-UK VOC B.1.1.7 sequences intermix within those from the UK (**Figure 2**). As convergent evolution can induce incorrect clustering[12], the same approach was repeated after excluding variable positions that define the

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.12.426373; this version posted January 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

B.1.1.7. lineage (**Supplementary Table 2**), which yielded a similar picture. These patterns are in line with the view that this variant succesfully spread around the world after it arose in the UK.



WITHDRAWN
see manuscript DOI for details

**Figure 2. SARS-CoV-2 variant B.1.1.7 arose in the UK and spread globally from there.** Tips in the phylogeny are colored according to lineage and country of origin (blue denotes taxa from UK, red denotes taxa from outside UK, and black corresponds to other lineages). (*) The branch leading tot the VOC B.1.1.7 clade has close to perfect suppport (0.99 Shimodaira Hasegawa (SH) branch support[9-11]). **A.** All sequences included. **B.** Illustrative subset of B 1.1.7 taxa from UK (in blue) and other countries (in red). The branch length scale (s/s/y) is indicated at the bottom. Tree display was obtained with the R package "ggtree"[13].

To estimate the timing of introduction of B.1.1.7 variants outside the UK, we applied a multistep analytic approach as previously described by our group for HIV[14,15] (see Supplementary Information). We identified a total of 30 clades of size ≥2 for a total of 152 sequences (ranging from 2 to 21) including only B.1.1.7 variants from outside the UK. More than two-thirds (22/30) were European clusters (**Supplementary Table 1**).

The earliest estimated seeding of B.1.1.7 from the UK dates to September 23rd, and the most recent to December 23rd, see **Supplementary Table 3** and **Figure 3A**). The number of weekly

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.12.426373; this version posted January 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

introductions (**Figure 3B**) peaked on the week of November 16th, while the peak of detection was in mid-December.

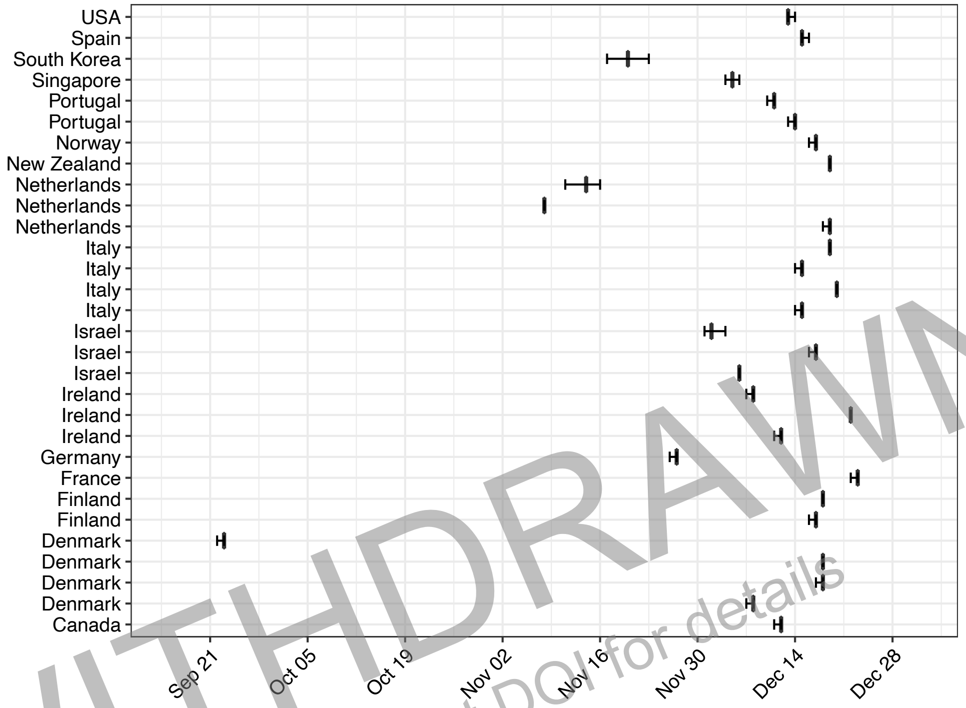
In response to the rapid increase in viral infections and spread, UK officials announced a lockdown on October 31st that came into force on November 5th and ended on December 5th. Given time to the most recent common ancestor (TMRCA) estimates, we determined that 20% (6/30) of the exportation events that gave rise to detectable non-UK VOC B1.1.7 transmission lineages occurred during this period (the remaining 80% occurred before or after these dates). The emergence and rapid dispersal of this new VOC led to the implementation of a new national strict lockdown on January 4, 2021[16].

As previously described by du Plessis *et al.*[17], we next used the TMRCA of each non-UK clade to estimate the genomic "detection lag" for each cluster, which represents the duration that a transmission lineage went undetected before it was first sampled by genome sequencing. The mean detection lag was ~10 days (IQR= 4-9.5). This largely agrees with detection lag-time estimates from SARS-CoV-2 importation *into* the UK in the first months of the pandemic[17], which was on average 8 days (IQR=3-15, ~10 days for lineages comprising ≤10 genomes and <1 day for lineages of >100 genomes).

Of note, virus genome sequences have been determined for only a fraction of infections. Even in the UK, where the by far largest sequencing effort is done, only an estimated 4.3% (129,939 available sequences out of 3,039,797 cases reported on January 7th)[18] of infections have been sequenced. For this reason, and also because not all sequenced SARS-CoV-2 genomes are being deposited in the GISAID repository, many B.1.1.7 variants that succesfully established transmission chains outside of the UK likely remain undetected (for now). Our estimated number of B.1.1.7 exportation events from the UK thus represents an underestimate. The sparse sampling and sequencing also poses limits to the accuracy with which introduction events can be dated (see du Plessis and colleagues[19] for a more detailed explanation).

WITHDRAWN
see manuscript DOI for details

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.12.426373; this version posted January 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.



WITHDRAWN
see manuscript DOI for details

**Figure 3A**. **Timing of introduction of each non-UK VOC B.1.1.7**. For each non-UK VOC lineage, we estimated the timing of introduction by performing molecular clock estimation on 100 replicates based on the clock rate distribution from Plessis *et al*[19] (see Supplementary Information). Each horizontal bar represent a non-UK cluster. Mean, lower and upper 95% CI are shown. Country of origin of these clusters is indicated on the y axis.

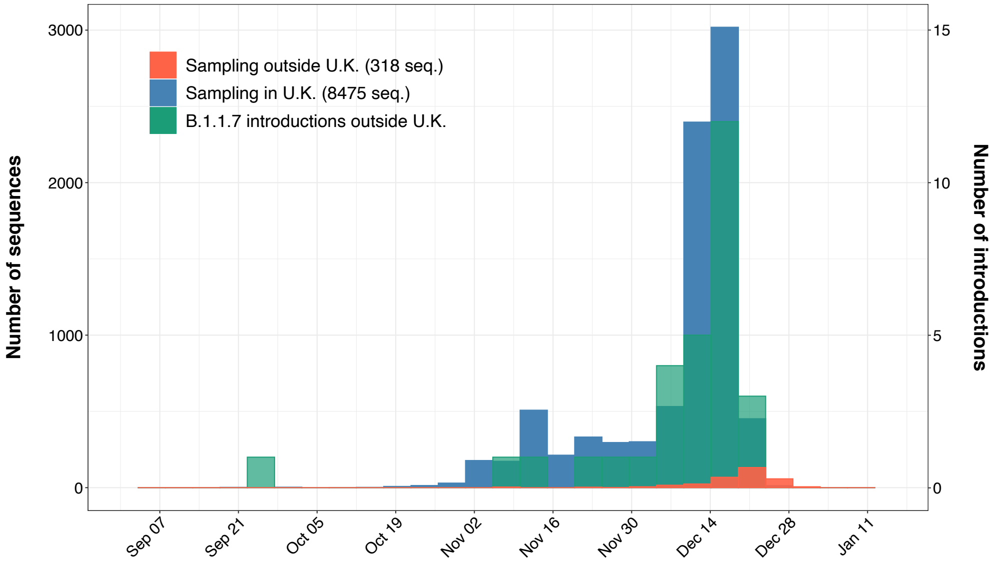


**Figure 3B. Number of introduction of B.1.1.7 outside UK.** Vertical green bar represents the biweekly number of introduction (right y axis). Red bars represent the number of B1.1.7 sequences collected through time.

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.12.426373; this version posted January 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Our results do not suggest that the canonical mutations of VOC B.1.1.7 evolved independently in different locations. Instead, our analyses point to an origin in and spread of the VOC B.1.1.7 from the UK. As for the virus' initial[20] and subsequent[21,22] spread, global connectedness and high levels of human mobility undoubtedly facilitated VOC B.1.1.7 dissemination. The swift global spread of VOC B.1.1.7 illustrates that current restrictions are insufficient to prevent the spread of new and emerging variants[23-29]. Similar to Ebola[30], HCV[31,32] and HIV[15], countermeasures to SARS-CoV-2 spread should be developed with a broader perspective than the national level. Otherwise, without population immunity, successful local reductions in SARS-CoV-2 burden will be counteracted by imported infections that set off new waves of viral spread, possibly exacerbated by novel phenotypic characteristics of the imported strains.

WITHDRAWN
see manuscript DOI for details

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.12.426373; this version posted January 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

## Supplementary Information

*Data collection and preparation.* All publicity available full-length SARS-CoV2 genomic sequences were collected from GISAID on January 7, 2020. Sequences were aligned using MAFFT and highly homoplasic sites were masked[33]. To reduce the data set size while maintaining an appropriate set of epidemiologically relevant background sequences, we used BLAST[34,35] to identify the 50 closest non-B.1.1.7 variants to each of the 8,786 B.1.1.7 genomic sequences in the data set[17,36]. After keeping one copy of duplicated entries that ranked among the 50 best hits, a total of 3,163 sequences out of the 284,666 non-B1.1.7 sequences available on GISAID were kept for further analyses and combined with the B.1.1.7 dataset.

*Identification of non-UK B.1.1.7 clades*. From a Maximum Likelihood (ML) phylogeny inferred using IQ-TREE v 2.1.2[8], B.1.1.7 clusters of size ≥2 including only non-UK sequences were identified in R[37].

*Timing of introduction*. For each non-UK clade, the phylogeny was rescaled into units of time with treedater[13], assuming a strict molecular clock with the rate of SARS-CoV-2 genome evolution drawn from an externally-estimated distribution as described by du Plessis *et* al[19]. Specifically, for the rate a normal distribution was specified with mean $9.41\times10^{-4}$ nucleotide substitutions per site per year and a standard deviation of $4.99\times10^{-5}$. To incorporate uncertainty in the estimated clock rate, molecular clock estimation was replicated 100 times for each non-UK B.1.1.7 clade.

WITHDRAWN
see manuscript DOI for details

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.12.426373; this version posted January 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

## Supplementary Tables.

### Supplementary table 1. Sampling distribution of B.1.17 sequences.

| Country | Count (n) | Percentage (%) |
|---|---|---|
| Australia | 8 | 0.091 |
| Brazil | 2 | 0.023 |
| Canada | 5 | 0.057 |
| Denmark | 74 | 0.842 |
| Finland | 14 | 0.159 |
| France | 14 | 0.159 |
| Germany | 5 | 0.057 |
| Gibraltar | 1 | 0.011 |
| Hong Kong | 3 | 0.034 |
| India | 3 | 0.034 |
| Ireland | 11 | 0.125 |
| Israel | 25 | 0.285 |
| Italy | 26 | 0.296 |
| Japan | 19 | 0.216 |
| Luxembourg | 3 | 0.034 |
| Netherlands | 22 | 0.25 |
| New Zealand | 6 | 0.068 |
| Norway | 5 | 0.057 |
| Oman | 1 | 0.011 |
| Pakistan | 2 | 0.023 |
| Portugal | 31 | 0.353 |
| Singapore | 6 | 0.068 |
| South Korea | 3 | 0.034 |
| Spain | 4 | 0.046 |
| Sweden | 10 | 0.114 |
| Switzerland | 6 | 0.068 |
| **United Kingdom** | **8468** | **96.381** |
| **USA** | **8** | **0.091** |
| Vietnam | 1 | 0.011 |
| **Total** | **8786** | **100** |

WITHDRAWN
see manuscript DOI for details

**Supplementary table 2.** Non-synonymous mutations and deletions to occur on the phylogenetic branch leading to lineage B.1.1.7.[1]

| gene | nucleotide | amino acid |
|---|---|---|
| ORF1ab | C3267T | T1001I |
|  | C5388A | A1708D |
|  | T6954C | I2230T |
|  | 11288-11296 deletion | SGF 3675-3677 deletion |
| spike | 21765-21770 deletion | HV 69-70 deletion |
|  | 21991-21993 deletion | Y144 deletion |
|  | A23063T | N501Y |
|  | C23271A | A570D |
|  | C23604A | P681H |
|  | C23709T | T716I |
|  | T24506G | S982A |
|  | G24914C | D1118H |
| Orf8 | C27972T | Q27stop |
|  | G28048T | R52I |
|  | A28111G | Y73C |
| N | 28280 GAT->CTA | D3L |
|  | C28977T | S235F |

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.12.426373; this version posted January 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

**Supplementary table 3**. Characteristics of the non-UK clusters identified.

| **continent** | **country (region)** | **cluster size** | **Estimated time of introduction, mean [95%CI]** | **detection lag in days [95%CI]** |
|---|---|---|---|---|
| Asia | Israel (NA) | 19 | 2020-12-02 [2020-12-01 - 2020-12-04] | 23 |
|  | Israel (NA) | 2 | 2020-12-06 [2020-12-06 - 2020-12-06] | 11 |
|  | Israel (NA) | 2 | 2020-12-17 [2020-12-16 - 2020-12-17] | 4 |
|  | Singapore (NA) | 3 | 2020-12-05 [2020-12-04 - 2020-12-06] | 15 |
|  | South Korea (Southkorea) | 3 | 2020-11-20 [2020-11-17 - 2020-11-23] | 33 |
| Europe | Denmark (Nordjylland) | 64 | 2020-09-23 [2020-09-22 - 2020-09-23] | 48 |
|  | Denmark (Sjaelland) | 2 | 2020-12-18 [2020-12-18 - 2020-12-18] | 4 |
|  | Denmark (Hovedstaden) | 2 | 2020-12-08 [2020-12-07 - 2020-12-08] | 7 |
|  | Denmark (Nordjylland) | 2 | 2020-12-18 [2020-12-17 - 2020-12-18] | 4 |
|  | Finland (Uusimaa) | 3 | 2020-12-18 [2020-12-18 - 2020-12-18] | 5 |
|  | Finland (Uusimaa) | 2 | 2020-12-17 [2020-12-16 - 2020-12-17] | 6 |
|  | France (Nouvelle-Aquitaine) | 2 | 2020-12-23 [2020-12-22 - 2020-12-23] | 4 |
|  | Germany (Lowersaxony) | 2 | 2020-11-27 [2020-11-26 - 2020-11-27] | 4 |
|  | Ireland (Wexford) | 3 | 2020-12-08 [2020-12-07 - 2020-12-08] | 10 |
|  | Ireland (Dublin) | 2 | 2020-12-22 [2020-12-22 - 2020-12-22] | 0 |
|  | Ireland (Dublin) | 2 | 2020-12-12 [2020-12-11 - 2020-12-12] | 7 |
|  | Italy (Campania) | 4 | 2020-12-19 [2020-12-19 - 2020-12-19] | 2 |
|  | Italy (Abruzzo) | 2 | 2020-12-15 [2020-12-14 - 2020-12-15] | 4 |
|  | Italy (Marche) | 2 | 2020-12-15 [2020-12-14 - 2020-12-15] | 4 |
|  | Italy (Campania) | 2 | 2020-12-20 [2020-12-20 - 2020-12-20] | 1 |
|  | Netherlands (Noord-Holland) | 8 | 2020-11-14 [2020-11-11 - 2020-11-16] | 30 |
|  | Netherlands (Noord-Holland) | 2 | 2020-11-08 [2020-11-08 - 2020-11-08] | 22 |
|  | Netherlands (Gelderland) | 2 | 2020-12-19 [2020-12-18 - 2020-12-19] | 4 |
|  | Norway (Vestland) | 2 | 2020-12-17 [2020-12-16 - 2020-12-17] | 6 |
|  | Portugal (NA) | 2 | 2020-12-14 [2020-12-13 - 2020-12-14] | 4 |
|  | Portugal (NA) | 2 | 2020-12-11 [2020-12-10 - 2020-12-11] | 5 |
|  | Spain (Madrid) | 2 | 2020-12-15 [2020-12-15 - 2020-12-16] | 7 |
| North America | Canada (Ontario) | 2 | 2020-12-12 [2020-12-11 - 2020-12-12] | 4 |
|  | USA (California) | 3 | 2020-12-13 [2020-12-13 - 2020-12-14] | 8 |
| Oceania | New Zealand (Auckland) | 2 | 2020-12-19 [2020-12-19 - 2020-12-19] | 1 |
| **Summary** |  | **152** | **2020-12-08 [2020-09-23 - 2020-12-23]** | **10 [I1 - 37]** |

WITHDRAWN
see manuscript DOI for details

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.12.426373; this version posted January 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

**Acknowledgements.** We acknowledge the important work of SARS-CoV-2 genome data producers globally contributing sequence data to the GISAID database. This work was supported by grants from the NIH (San Diego Center for AIDS Research, CFAR). AC was supported by NIH Grant AI131971 (R21). BV was supported by a postdoctoral grant (12U7121N) of the Research Foundation -- Flanders (Fonds voor Wetenschappelijk Onderzoek). SD is supported by the Fonds National de la Recherche Scientifique (FNRS, Belgium). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

WITHDRAWN
see manuscript DOI for details

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.12.426373; this version posted January 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

WITHDRAWN
see manuscript DOI for details

## REFERENCES

1 Rambaut, A. *et al.* Preliminary genomic characterisation of an emergent SARS-CoV-2 lineage in the UK defined by a novel set of spike mutations. *virological.org* (2020).

2 Volz, E. *et al.* Transmission of SARS-CoV-2 Lineage B.1.1.7 in England: Insights from linking epidemiological and genetic data. *medRxiv*, 2020.2012.2030.20249034, doi:10.1101/2020.12.30.20249034 (2021).

3 Chand, M. *et al.* Variant of Concern 202012/01. *Public Health England* (2020).

4 Russell, G., Woodhouse, A., Dempsey, H., Clarfelt, H. & Ralph, O. Coronavirus: New York detects first case of UK variant, California reveals further occurrences — as it happened. *Financial Times* (2020).

5 Volz, E. *et al.* Evaluating the Effects of SARS-CoV-2 Spike Mutation D614G on Transmissibility and Pathogenicity. *Cell* **184**, 64-75.e11, doi:10.1016/j.cell.2020.11.020 (2021).

6 Shu, Y. & McCauley, J. GISAID: Global initiative on sharing all influenza data - from vision to reality. *Euro Surveill* **22**, doi:10.2807/1560-7917.Es.2017.22.13.30494 (2017).

7 Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular biology and evolution* **30**, 772-780, doi:10.1093/molbev/mst010 (2013).

8 Minh, B. Q. *et al.* IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. *Mol Biol Evol* **37**, 1530-1534, doi:10.1093/molbev/msaa015 (2020).

9 Guindon, S. *et al.* New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0. *Syst Biol* **59**, 307-321, doi:10.1093/sysbio/syq010 (2010).

10 Guindon, S. & Gascuel, O. A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood. *Systematic biology* **52**, 696-704 (2003).

11 Shimodaira, H. & Hasegawa, M. Multiple Comparisons of Log-Likelihoods with Applications to Phylogenetic Inference. *Molecular Biology and Evolution* **16**, 1114-1114, doi:10.1093/oxfordjournals.molbev.a026201 (1999).

12 Lemey, P. *et al.* Molecular footprint of drug-selective pressure in a human immunodeficiency virus transmission chain. *Journal of virology* **79**, 11981-11989, doi:10.1128/JVI.79.18.11981-11989.2005 (2005).

13 Volz, E. M. & Frost, S. D. W. Scalable relaxed clock phylogenetic dating. *Virus Evolution* **3**, doi:10.1093/ve/vex025 (2017).

14 Vrancken, B. *et al.* Comparative Circulation Dynamics of the Five Main HIV Types in China. *J Virol* **94**, doi:10.1128/jvi.00683-20 (2020).

15 Vrancken, B. *et al.* Dynamics and Dispersal of Local HIV Epidemics Within San Diego and Across The San Diego-Tijuana Border. *Clin Infect Dis*, doi:10.1093/cid/ciaa1588 (2020).

16 Steed, L. & Cavanagh, N. LOCKED IN When did lockdown start in the UK? *The Sun* (2021).

17 Vrancken, B. *et al.* The multi-faceted dynamics of HIV-1 transmission in Northern Alberta: A combined analysis of virus genetic and public health data. *Infection, genetics and evolution : journal of molecular epidemiology and evolutionary genetics in infectious diseases* **52**, 100-105, doi:10.1016/j.meegid.2017.04.005 (2017).

18 GOV.UK. *Coronavirus (COVID-19) in the UK.*, <https://coronavirus.data.gov.uk/details/cases> (2020).

19 du Plessis, L. *et al.* Establishment &amp;amp; lineage dynamics of the SARS-CoV-2 epidemic in the UK. *medRxiv*, 2020.2010.2023.20218446, doi:10.1101/2020.10.23.20218446 (2020).

20 Worobey, M. *et al.* The emergence of SARS-CoV-2 in Europe and the US. *bioRxiv : the preprint server for biology*, 2020.2005.2021.109322, doi:10.1101/2020.05.21.109322 (2020).

21 Badr, H. S. & Gardner, L. M. Limitations of using mobile phone data to model COVID-19 transmission in the USA. *Lancet Infect Dis*, doi:10.1016/s1473-3099(20)30861-6 (2020).

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.12.426373; this version posted January 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

22 Badr, H. S. *et al.* Association between mobility patterns and COVID-19 transmission in the USA: a mathematical modelling study. *Lancet Infect Dis* **20**, 1247-1254, doi:10.1016/s1473-3099(20)30553-3 (2020).

23 Peiris, J. S., Yuen, K. Y., Osterhaus, A. D. & Stöhr, K. The severe acute respiratory syndrome. *N Engl J Med* **349**, 2431-2441, doi:10.1056/NEJMra032498 (2003).

24 Johnson, N. P. & Mueller, J. Updating the accounts: global mortality of the 1918-1920 "Spanish" influenza pandemic. *Bull Hist Med* **76**, 105-115, doi:10.1353/bhm.2002.0022 (2002).

25 Shortridge, K. F., Peiris, J. S. & Guan, Y. The next influenza pandemic: lessons from Hong Kong. *J Appl Microbiol* **94 Suppl**, 70s-79s, doi:10.1046/j.1365-2672.94.s1.8.x (2003).

26 Subbarao, K. & Katz, J. Avian influenza viruses infecting humans. *Cellular and Molecular Life Sciences CMLS* **57**, 1770-1784 (2000).

27 Chua, K. B. *et al.* Nipah virus: a recently emergent deadly paramyxovirus. *Science* **288**, 1432-1435, doi:10.1126/science.288.5470.1432 (2000).

28 Nash, D. *et al.* The outbreak of West Nile virus infection in the New York City area in 1999. *N Engl J Med* **344**, 1807-1814, doi:10.1056/nejm200106143442401 (2001).

29 Reid, A. H. & Taubenberger, J. K. The origin of the 1918 pandemic influenza virus: a continuing enigma. *J Gen Virol* **84**, 2285-2292, doi:10.1099/vir.0.19302-0 (2003).

30 Kamorudeen, R. T., Adedokun, K. A. & Olarinmoye, A. O. Ebola outbreak in West Africa, 2014 - 2016: Epidemic timeline, differential diagnoses, determining factors, and lessons for future response. *J Infect Public Health* **13**, 956-962, doi:10.1016/j.jiph.2020.03.014 (2020).

31 Pérez, A. B. *et al.* Increasing importance of European lineages in seeding the hepatitis C virus subtype 1a epidemic in Spain. *Euro Surveill* **24**, doi:10.2807/1560-7917.Es.2019.24.9.1800227 (2019).

32 Vrancken, B. *et al.* Cross-country migration linked to people who inject drugs challenges the long-term impact of national HCV elimination programmes. *J Hepatol* **71**, 1270-1272, doi:10.1016/j.jhep.2019.08.010 (2019).

33 De Maio, N. *et al.* Issues with SARS-CoV-2 sequencing data. *virological.org* (2020).

34 Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *Journal of molecular biology* **215**, 403-410, doi:10.1016/S0022-2836(05)80360-2 (1990).

35 Camacho, C. *et al.* BLAST+: architecture and applications. *BMC Bioinformatics* **10**, 421, doi:10.1186/1471-2105-10-421 (2009).

36 Gräf, T. *et al.* Contribution of Epidemiological Predictors in Unraveling the Phylogeographic History of HIV-1 Subtype C in Brazil. *J Virol* **89**, 12341-12348, doi:10.1128/jvi.01681-15 (2015).

37 Paradis, E. & Schliep, K. ape 5.0: an environment for modern phylogenetics and evolutionary analyses in R. *Bioinformatics* **35**, 526-528, doi:10.1093/bioinformatics/bty633 (2019).

WITHDRAWN
see manuscript DOI for details

bioRxiv preprint doi: https://doi.org/10.1101/2021.01.12.426373; this version posted January 13, 2021. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

WITHDRAWN
see manuscript DOI for details