nature

https://doi.org/10.1038/s41586-021-03677-y

Accelerated Article Preview

# Spread of a SARS-CoV-2 variant through Europe in the summer of 2020

Received: 25 November 2020

Accepted: 28 May 2021

Accelerated Article Preview Published online 7 June 2021

Cite this article as: Hodcroft, E. B. et al. Spread of a SARS-CoV-2 variant through Europe in the summer of 2020. *Nature* https://doi.org/10.1038/s41586-021-03677-y (2021).

**Emma B. Hodcroft, Moira Zuber, Sarah Nadeau, Timothy G. Vaughan, Katharine H. D. Crawford, Christian L. Althaus, Martina L. Reichmuth, John E. Bowen, Alexandra C. Walls, Davide Corti, Jesse D. Bloom, David Veesler, David Mateo, Alberto Hernando, Iñaki Comas, Fernando González Candelas, SeqCOVID-SPAIN consortium, Tanja Stadler & Richard A. Neher**

Article

# Spread of a SARS-CoV-2 variant through Europe in the summer of 2020

https://doi.org/10.1038/s41586-021-03677-y

Received: 25 November 2020

Accepted: 28 May 2021

Published online: 7 June 2021

Emma B. Hodcroft[1,2,3]✉, Moira Zuber[1], Sarah Nadeau[2,4], Timothy G. Vaughan[2,4], Katharine H. D. Crawford[5,6,7], Christian L. Althaus[3], Martina L. Reichmuth[3], John E. Bowen[8], Alexandra C. Walls[8], Davide Corti[9], Jesse D. Bloom[5,6,10], David Veesler[8], David Mateo[11], Alberto Hernando[11], Iñaki Comas[12,13,14], Fernando González Candelas[13,14,15], SeqCOVID-SPAIN consortium*, Tanja Stadler[2,4,93] & Richard A. Neher[1,2,93]✉

Following its emergence in late 2019, the spread of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2)[1,2] has been tracked via phylogenetic analysis of viral genome sequences in unprecedented detail[3–5]. While the virus spread globally in early 2020 before borders closed, intercontinental travel has since been greatly reduced. However, within Europe travel resumed in the summer of 2020. Here we report on a novel SARS-CoV-2 variant, 20E (EU1), that emerged in Spain in early summer, and subsequently spread across Europe. We find no evidence of increased transmissibility, but instead demonstrate how rising incidence in Spain, resumption of travel, and lack of effective screening and containment may explain the variant's success. Despite travel restrictions, we estimate 20E (EU1) was introduced hundreds of times to European countries by summertime travelers, likely undermining local efforts to keep SARS-CoV-2 cases low. Our results demonstrate how a variant can rapidly become dominant even in absence of a substantial transmission advantage in favorable epidemiological settings. Genomic surveillance is critical to understanding how travel can impact SARS-CoV-2 transmission, and thus for informing future containment strategies as travel resumes.

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) is the first pandemic where the spread of a viral pathogen has been globally tracked in near real-time using phylogenetic analysis of viral genome sequences[3–5]. SARS-CoV-2 genomes continue to be generated at a rate far greater than for any other pathogen and more than 950,000 full genomes are available on GISAID as of April 2021[6].

In addition to tracking the viral spread, these sequences have been used to monitor mutations which might change the transmission, pathogenesis, or antigenic properties of the virus. One mutation in particular, D614G in the spike protein (Nextstrain clade 20A and its descendants), seeded large outbreaks in Europe in early 2020 and subsequently dominated the outbreaks in the Americas, thereby largely replacing previously circulating lineages. This rapid rise led to the suggestion that this variant is more transmissible, which has since been corroborated by phylogenetic[7,8] and experimental evidence[9,10]. Subsequently, three variants of concern (VoCs), 501Y.V1/B.1.1.7[11,12], 501Y.V2/B.1.351[13,14] and 501Y.V3/P.1[15] with increased transmissibility and/or partial neutralization escape, were identified at the end of 2020.

Following the global dissemination of SARS-CoV-2 in early 2020[3], intercontinental travel dropped dramatically. Within Europe, however, travel and in particular holiday travel resumed in summer. Here we report on a SARS-CoV-2 variant 20E (EU1) (S:A222V) that emerged in early summer 2020, presumably in Spain, and subsequently spread to multiple locations in Europe, rising in frequency in parallel. As we report here, this variant, 20E (EU1), and a second variant 20A.EU2 with mutation S477N in the spike protein accounted for the majority of sequences in Europe in the autumn of 2020.

## European variants in Summer 2020

Figure 1 shows a time scaled phylogeny of sequences sampled in Europe through the end of November and their global context, highlighting the variants discussed here. A cluster of sequences in clade 20A has an additional mutation S:A222V colored in orange. We designate this cluster as 20E (EU1) (this cluster consists of lineage B.1.177 and its sublineages[16]).

In addition to 20E (EU1), a variant (20A.EU2; blue in Fig. 1) with several amino acid substitutions, including S:S477N, became common in

[1]Biozentrum, University of Basel, Basel, Switzerland. [2]Swiss Institute of Bioinformatics, Basel, Switzerland. [3]Institute of Social and Preventive Medicine, University of Bern, Bern, Switzerland. [4]Department of Biosystems Science and Engineering, ETH Zürich, Basel, Switzerland. [5]Division of Basic Sciences and Computational Biology Program, Fred Hutchinson Cancer Research Center, Seattle, WA, 98109, USA. [6]Department of Genome Sciences, University of Washington, Seattle, WA, 98195, USA. [7]Medical Scientist Training Program, University of Washington, Seattle, WA, 98195, USA. [8]Department of Biochemistry, University of Washington, Seattle, WA, USA. [9]Humabs Biomed SA, a subsidiary of Vir Biotechnology, 6500, Bellinzona, Switzerland. [10]Howard Hughes Medical Institute, Seattle, WA, 98103, USA. [11]Kido Dynamics SA, Avenue de Sevelin 46, 1004, Lausanne, Switzerland. [12]Tuberculosis Genomics Unit, Biomedicine Institute of Valencia (IBV-CSIC), Valencia, Spain. [13]CIBER de Epidemiología y Salud Pública (CIBERESP), Madrid, Spain. [14]on behalf or the SeqCOVID-SPAIN consortium, Valencia, Spain. [15]Joint Research Unit "Infection and Public Health" FISABIO-University of Valencia, Institute for Integrative Systems Biology (I2SysBio), Valencia, Spain. [93]These authors contributed equally: Tanja Stadler, Richard A. Neher. *A list of authors and their affiliations appears at the end of the paper. ✉e-mail: emma.hodcroft@ispm.unibe.ch; richard.neher@unibas.ch

some European countries, particularly France (Fig. ED1). The S:S477N substitution has arisen multiple times independently, for example in clade 20F that dominated the outbreak in Oceania during the southern-hemisphere winter. Residue S477 is close to the receptor binding site (Fig ED2) and part of the epitope recognized by the S2E12 and C102 neutralizing antibodies[17,18].

Several other smaller clusters defined by the spike mutations D80Y, S98F, N439K are also seen in multiple countries (see Table ED I and Fig. ED1). While none of these have reached the prevalence of 20E (EU1) or 20A.EU2, some have attracted attention in their own right: S:N439K is present in two larger clusters found across Europe[19] and arose several times independently. Updated phylogenies and further analyses for these and other variants are available at CoVariants.org.

## Characterization of S:A222V

Our analysis here focuses on the variant 20E (EU1) with substitution S:A222V in the spike protein's domain A (Fig. ED2) also referred to as the N-terminal domain (NTD)[18,20,21]. S:A222V is not known to play a direct role in receptor binding or membrane fusion for SARS-CoV-2. However, mutations can sometimes mediate long-range effects on protein conformation or stability.

To evaluate if the A222V mutation affects the conformation of the SARS-CoV-2 spike glycoprotein, we probed binding of the benchmark COVID-19 convalescent patient plasma from the National Institute for Biologicals Standards and Control, and neutralizing monoclonal antibodies recognizing the RBD (S2E12 and S309[18,22,23]) and NTD (4A8)[24]. The dose-response curves were indistinguishable for the SARS-CoV-2 2PS and the SARS-CoV-2 2P A222V D614G S ectodomain trimers (observed by ELISA, Fig. ED3a-d), aligning with results from a recent study[25]. Collectively, these data indicate that the A222V substitution does not affect the SARS-CoV-2 S antigenicity appreciably.

To test whether the A222V mutation had an obvious functional effect on spike's ability to mediate viral entry, we produced lentiviral particles pseudotyped with spike either containing or lacking the A222V mutation in the background of the D614G mutation and deletion of the end of spike's cytoplasmic tail. Lentiviral particles with the A222V mutant spike had slightly higher titers than those without (mean 1.3-fold higher), although the difference was not statistically significant after normalization by p24 concentration (Fig. ED3e-h). Therefore, A222V does not lead to the same large increases in the titers of spike-pseudotyped lentivirus that has been observed for the D614G mutation[7,10] However, this small effect must be interpreted cautiously, as the effects of mutations on actual viral transmission in humans are not always paralleled by measurements made in simplified experimental systems.

In addition to S:A222V, 20E (EU1) has the amino acid mutations *ORF10*:V30L, *N*:A220V and *ORF14*:L67F. However, there is little evidence of the functional relevance of *ORF10* and *ORF14*[26,27]. Different mutations between positions 180 and 220 in *N* are observed in almost every major lineage and we are not aware of any evidence suggesting that these mutations have important phenotypic consequence. Therefore, we examined epidemiological and phylogenetic evidence to explain the spread of 20E (EU1).

## Early observations of 20E (EU1)

The earliest sequences were sampled on the 20th of June, (7 Spanish and 1 Dutch sequence). By the end of August, 20E (EU1) also included sequences from Belgium, Switzerland, France, Denmark, the UK, Germany, Latvia, Sweden, Norway and Italy. Sequences from Hong Kong, Australia, New Zealand, and Singapore, presumably exports from Europe, were first detected between mid-August and mid-October (see Supplementary Table I).

The proportion of sequences from several countries which fall into 20E (EU1), by ISO week, is plotted in Fig. 2. 20E (EU1) first rose in frequency in Spain, jumping to around 50% prevalence within a month of the first sequence being detected before rising to 80%. In many European countries, we observe a gradual rise starting in mid-July before settling at a level between 15 and 80% in September or October.

## Expansion and spread across Europe

To quantify the spread of EU1 across Europe, we constructed a phylogeny (Fig. ED4a) based on data from samples collected before 2020-09-30 and available on GISAID in Jan 2021, as described in Methods. The phylogeny is collapsed to group diversity possibly stemming from within-country transmission into sectors of the pie-charts (see Fig. ED4b-d) for selected countries. The tree indicates that 20E (EU1) harbors substantial diversity and most major genotypes have been observed in many European countries. Since it is unlikely that phylogenetic patterns sampled in multiple countries arose independently, it is reasonable to assume that the majority of mutations observed in the tree arose once and were carried (possibly multiple times) between countries. Throughout July and August 2020, Spain had a higher per capita incidence than most other European countries (see Fig ED5) and 20E (EU1) was much more prevalent in Spain then elsewhere, suggesting Spain as likely origin of most 20E (EU1) introductions to other countries.

Epidemiological data from Spain indicates the earliest sequences in the cluster are associated with two known outbreaks in the north-east of the country. The variant seems to have initially spread among agricultural workers in Aragon and Catalonia, then moved into the local population, where it was able to travel to the Valencia Region and on to the rest of the country.

Most basal genotypes have been observed both in Spain and a large number of other countries, suggesting repeated exports. However, the 795 sequences from Spain contributing to Fig. ED4a likely do not represent the full diversity. Variants found only outside of Spain may reflect diversity that arose in secondary countries, or may represent diversity present but not sampled in Spain (particularly as some European countries like the UK and Denmark sequence a high proportion of cases). Despite limitations in sampling, Fig. ED4a clearly shows that most major genotypes in this cluster were distributed to multiple countries, suggesting that identical genotypes were introduced into many countries . This is consistent with the large number of introductions estimated from travel data, discussed below. While initial introductions of the variant likely originated from Spain, 20E (EU1) cases outside of Spain surpassed those in Spain in late September and later cross-border transmissions likely originated in other countries (see Fig ED5 B). (See supplementary text for a discussion of travel restrictions in selected European countries and the associated patterns of 20E (EU1) introductions.)

Fig. ED4e shows the distribution of sequence clusters compatible with onward transmission within countries outside of Spain, highlighting two different patterns. Norway and Iceland, for example, seem to have only a small number of introductions over the summer that led to substantial further spread. In Fig. ED4a, the majority of sequences from these countries fall into one sector, the remainder are singletons or very small clusters that have not spread. However, later sequences in Norway or Iceland often cluster more closely with diversity in non-Spanish European countries, which may suggest further introductions came from third countries (see 20E (EU1) Nextstrain build online).

In contrast, countries like Switzerland, the Netherlands, or the United Kingdom have sampled sequences that correspond to a large number of independent introductions that include most major genotypes observed in Spain.

## No evidence for transmission advantage

During a dynamic outbreak, it is particularly difficult to unambiguously tell whether a particular variant is increasing in frequency because it

has an intrinsic advantage, or because of epidemiological factors[28]. In fact, it is a tautology that every novel big cluster must have grown recently and multiple lines of independent evidence are required in support of an intrinsically elevated transmission potential.

20E (EU1) was dispersed across Europe initially mainly by travelers to and from Spain. Many EU and Schengen-area countries opened their borders to other countries in the bloc on 15th June. Travel resumed quickly and peaked during July and August, see Fig. 3. The number of confirmed SARS-CoV-2 cases in Spain rose from around 10 cases per 100k inhabitants per week in early July to 100 in late August, while case numbers remained low in most of Europe during this time. To explore whether repeated imports are sufficient to explain the rapid rise in frequency and the displacement of other variants, we first estimated the number of expected introductions of 20E (EU1) based on the number of visitors from a particular country to different provinces of Spain and the SARS-CoV-2 incidence in the provinces. Taking reported incidence in the provinces at face value and assuming that returning tourists have a similar incidence, we expect 380 introductions of 20E (EU1) into the UK over the summer (6 July-27 Sept, see Supplementary Table II and Fig. 3 for tourism summaries[29] and departure statistics[30]). Similarly, for Germany and Switzerland we would expect around 320 and 90 introductions of 20E (EU1), respectively. We then create a simple model that also incorporates the incidence in the country where travelers are returning to and onward spread of imported 20E (EU1) cases to estimate the frequency of 20E (EU1) in countries across Europe over time (see Fig. 3). This model assumes that 20E (EU1) spread at the same rate as other variants in the resident countries and predicts that the frequencies of 20E (EU1) would start rising in July, continue to rise through August, and be stable thereafter in concordance with observations in many countries (see Fig. 3 B).

While the shape of the expected frequency trajectories from imports in Fig. 3 B is consistent with observations, this naive import model underestimates the final observed frequency of 20E (EU1) by between 1- and 12-fold depending on the country, see Fig. ED6. This discrepancy might be due to either intrinsically faster transmission of 20E (EU1) or due to underestimation of introductions. Underestimates might be due to country-specific reporting such as the relative ascertainment rate in source and destination populations and the fact that risk of exposure and onward transmission are likely increased by travel-related activities both abroad, en route, and at home. Furthermore, SARS-CoV-2 incidence in holiday destinations may not be well-represented by the provincial averages used in the model. For example, during the first wave in spring 2020, some ski resorts had exceptionally high incidence and contributed disproportionately to dispersal of SARS-CoV-2[31,32]. The fact that the rapid increase of the frequency of 20E (EU1) slowed or stopped in most countries after the summer travel period and didn't fully replace other variants is consistent with import driven dynamics with little or no competitive advantage.

The notion that an underestimated incidence in travel returnees rather than faster spread of 20E (EU1) is the major contributor to above discrepancy is supported by the fact that German authorities report about 2.2 times as many cases with suspected infection in Spain than the model predicts (982 reported vs 452 estimated from 6 July-13 Sept regardless of variant), see Fig. ED7 A. Switzerland reported 131 infections in travel returnees, while the model predicts 130. After adjusting imports for the 37% of Swiss case reports without exposure information, the model underestimates introductions 1.6-fold. Countries with small (1-4 fold) and large (8-12 fold) discrepancies tend to visit distinct destinations in Spain, see Figs. ED6 and ED7(c-e), further suggesting that underestimation of incidence in travel returnees is determined by destination and behavior.

To investigate the possibility of faster growth of 20E (EU1) introductions, we identified 20E (EU1) and non-20E (EU1) introductions into Switzerland and their downstream Swiss transmission chains. These data suggest 34 or 291 introductions of 20E (EU1) depending on the criterion used to assign sequences to putative transmission chains (see Methods). Phylodynamic estimates of the effective reproductive number ($R_e$) through time for introductions of 20E (EU1) and for other variants (see Fig. ED8) suggest a tendency for 20E (EU1) introductions to transiently grow faster. This transient signal of faster growth, however, is more readily explained with behavioral differences and increased travel-associated transmission than intrinsic differences to the virus. We repeated the phylodynamic analysis with a pan-European set of putative introductions showing similar patterns as observed for Switzerland.

These patterns are further consistent with the fact that Swiss cases with likely exposure in Spain tended to be in younger individuals (median 30 years, IQR 23-42.25 years) than cases acquired in Switzerland (median 35 years, IQR 24-51 years). These younger individuals tend to have more contacts than older age groups[33,34]. Such association with particular demographics will decay rapidly and with it any associated increased transmission inferred by phylodynamics.

Most 20E (EU1) introductions are expected to have occurred towards the end of summer when incidence in Spain was rising and return travel volume peaked. Comparatively high incidence of non-20E (EU1) variants at this time and hence a relatively low impact of imported variants (e.g. Belgium, see Fig. ED5) might explain why 20E (EU1) remains at low frequencies in some countries despite high-volume travel to Spain.

Case numbers across Europe started to rise rapidly around the same time the 20E (EU1) variant started to become prevalent in multiple countries (Fig. ED5). However, countries where 20E (EU1) was rare (Belgium, France, Czech Republic, Fig. ED1) have seen similarly rapid increases, suggesting that this rise was not driven by any particular lineage and that 20E (EU1) has no substantial difference in transmissibility. Furthermore, we observe in Switzerland that $R_e$ increased in fall by a comparable amount for the 20E (EU1) and non-20E (EU1) variants (see (Fig. ED8). While we cannot rule out that 20E (EU1) has a slight transmission advantage compared to other variants circulating at the time, most of its spread is explained by epidemiological factors., The arrival of fall and seasonal factors are a more plausible explanation for the resurgence of cases[35].

## Discussion

The rapid spread of 20E (EU1) and other variants underscores the importance of a coordinated and systematic sequencing effort to detect, track, and analyze emerging SARSCoV-2 variants. This becomes even more urgent with the recent detection of several VoCs[11–15]. It is only through multi-country genomic surveillance that it has been possible to detect and track 20E (EU1) and other variants.

When a new variant is observed, policy makers need a rapid assessment of whether the new variant increases the transmissibility of the virus, evades pre-existing immunity or has different clinical properties[36]. In case of 20E (EU1) none of these seem to have changed substantially, making it an important example of how travel combined with large regional differences in prevalence can lead to substantial rapid shifts in the variant distribution without a dramatic transmission advantage. Such shifts that are driven predominantly by epidemiological factors are more likely in a low incidence setting, where a large fraction of cases can be due to introductions. In contrast, the VoC 501Y.V1/B.1.1.7 spread across Europe in late 2020 while most countries, including the UK, where it first rose to prominence, reported high incidence. In such a high incidence setting, travel alone cannot explain a rapid rise in frequency and the dynamics points to a bona fide transmission advantage. In depth characterization of a spectrum of such dynamics (no substantial advantage in case of 20E (EU1), moderate advantage in case of D614G[8], and a strong transmission advantage of 501Y.V1/B.1.1.7[11,12] and 501Y.V2[13]) will facilitate assessment of emerging variants in the future.

Finally, our analysis highlights that countries should carefully consider their approach to travel when large-scale inter-country movement

resumes across Europe. We show that holiday travel in summer 2020 resulted in unexpectedly high levels of introductions and onward spread across Europe. Whether the 20E (EU1) variant described here has rapidly spread due to a transmission advantage or due to epidemiological factors alone, its repeated introduction and rise in prevalence in multiple countries implies that the summer travel guidelines and restrictions were generally not sufficient to prevent onward transmission of introductions. Travel precautions such as quarantine should in principle have prevented spread of SARS-CoV-2 infections acquired abroad, but in practice failed to have the desired effect. While long-term travel restrictions and border closures are not tenable or desirable, identifying better ways to reduce the risk of introducing variants, and ensuring that those which are introduced do not go on to spread widely, will help countries maintain often hard-won low levels of SARS-CoV-2 transmission.

## 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/s41586-021-03677-y.


1. WHO Emergency Committee. Statement on the second meeting of the International Health Regulations (2005) Emergency Committee regarding the outbreak of novel coronavirus (2019-nCoV). (2020).
2. Zhu, N. *et al*. Brief Report: A Novel Coronavirus from Patients with Pneumonia in China, *2019. N. Engl. J. Med*. **382**, 727 (2020).
3. Worobey, M. *et al*. The emergence of SARS-CoV-2 in Europe and North America. *Science* (2020) https://doi.org/10.1126/science.abc8169
4. Hadfield, J. *et al*. Nextstrain: real-time tracking of pathogen evolution. *Bioinformatics* (2018) https://doi.org/10.1093/bioinformatics/bty407.
5. Plessis, L. du *et al*. Establishment and lineage dynamics of the SARS-CoV-2 epidemic in the UK. *Science* **371**, 708–712 (2021).
6. Shu, Y. & McCauley, J. GISAID: Global initiative on sharing all influenza data – from vision to reality. *Eurosurveillance* **22**, 30494 (2017).
7. Korber, B. *et al*. Tracking Changes in SARS-CoV-2 Spike: Evidence that D614G Increases Infectivity of the COVID-19 Virus. *Cell* **182**, 812-827.e19 (2020).
8. Volz, E. *et al*. Evaluating the effects of SARS-CoV-2 Spike mutation D614G on transmissibility and pathogenicity. *Cell* **0**, (2020).
9. Plante, J. A. *et al*. Spike mutation D614G alters SARS-CoV-2 fitness. *Nature* 1–9 (2020) https://doi.org/10.1038/s41586-020-2895-3.
10. Yurkovetskiy, L. *et al*. Structural and Functional Analysis of the D614G SARS-CoV-2 Spike Protein Variant. *Cell* **183**, 739-751.e8 (2020).
11. Davies, N. G. *et al*. Estimated transmissibility and impact of SARS-CoV-2 lineage B.1.1.7 in England. *Science* (2021) https://doi.org/10.1126/science.abg3055.
12. Volz, E. *et al*. Assessing transmissibility of SARS-CoV-2 lineage B.1.1.7 in England. *Nature* 1–17 (2021) https://doi.org/10.1038/s41586-021-03470-x.
13. Pearson, C. A. B. *et al*. Estimates of severity and transmissibility of novel SARS-CoV-2 variant 501Y.V2 in South Africa. *CMMID Repos. Prepr.* (2021).
14. Tegally, H. *et al*. Detection of a SARS-CoV-2 variant of concern in South Africa. *Nature* 1–6 (2021) https://doi.org/10.1038/s41586-021-03402-9.
15. Sabino, E. C. *et al*. Resurgence of COVID-19 in Manaus, Brazil, despite high seroprevalence. *The Lancet* **397**, 452–455 (2021).
16. Rambaut, A. *et al*. A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. *Nat. Microbiol.* **5**, 1403–1407 (2020).
17. Barnes, C. O. *et al*. SARS-CoV-2 neutralizing antibody structures inform therapeutic strategies. *Nature* (2020) https://doi.org/10.1038/s41586-020-2852-1.
18. Tortorici, M. A. *et al*. Ultrapotent human antibodies protect against SARS-CoV-2 challenge via multiple mechanisms. *Science* **370**, 950–957 (2020).
19. Thomson, E. C. *et al*. The circulating SARS-CoV-2 spike variant N439K maintains fitness while evading antibody-mediated immunity. *bioRxiv* 2020.11.04.355842 (2020) https://doi.org/10.1101/2020.11.04.355842.
20. McCallum, M., Walls, A. C., Bowen, J. E., Corti, D. & Veesler, D. Structure-guided covalent stabilization of coronavirus spike glycoprotein trimers in the closed conformation. *Nat. Struct. Mol. Biol.* **27**, 942–949 (2020).
21. Walls, A. C. *et al*. Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. *Cell* **181**, 281-292.e6 (2020).
22. Pinto, D. *et al*. Cross-neutralization of SARS-CoV-2 by a human monoclonal SARS-CoV antibody. *Nature* **583**, 290–295 (2020).
23. Walls, A. C. *et al*. Elicitation of Potent Neutralizing Antibody Responses by Designed Protein Nanoparticle Vaccines for SARS-CoV-2. *Cell* **183**, 1367-1382.e17 (2020).
24. Chi, X. *et al*. A neutralizing human antibody binds to the N-terminal domain of the Spike protein of SARS-CoV-2. *Science* **369**, 650–655 (2020).
25. McCallum, M. *et al*. N-terminal domain antigenic mapping reveals a site of vulnerability for SARS-CoV-2. *bioRxiv* 2021.01.14.426475 (2021) https://doi.org/10.1101/2021.01.14.426475.
26. Finkel, Y. *et al*. The coding capacity of SARS-CoV-2. *Nature* 1–6 (2020) https://doi.org/10.1038/s41586-020-2739-1.
27. Pancer, K. *et al*. The SARS-CoV-2 ORF10 is not essential in vitro or in vivo in humans. *PLOS Pathog*. **16**, e1008959 (2020).
28. Grubaugh, N. D., Hanage, W. P. & Rasmussen, A. L. Making Sense of Mutation: What D614G Means for the COVID-19 Pandemic Remains Unclear. *Cell* **182**, 794–795 (2020).
29. Instituto Nacional de Estadistica. *Hotel Industry and Tourism – Tourist Movement on Borders Survey Frontur*. (2020).
30. Aena.es. *Air traffic statistics*. (2020).
31. Correa-Martinez, C. L. *et al*. A Pandemic in Times of Global Tourism: Superspreading and Exportation of COVID-19 Cases from a Ski Area in Austria. *J. Clin. Microbiol.* **58**, (2020).
32. Knabl, L. *et al*. High SARS-CoV-2 Seroprevalence in Children and Adults in the Austrian Ski Resort Ischgl. *medRxiv* 2020.08.20.20178533 (2020) https://doi.org/10.1101/2020.08.20.20178533.
33. Mossong, J. *et al*. Social Contacts and Mixing Patterns Relevant to the Spread of Infectious Diseases. *PLOS Med.* **5**, e74 (2008).
34. Jarvis, C. I. *et al*. Quantifying the impact of physical distance measures on the transmission of COVID-19 in the UK. *BMC Med.* **18**, 124 (2020).
35. Neher, R. A., Dyrdak, R., Druelle, V., Hodcroft, E. B. & Albert, J. Potential impact of seasonal forcing on a SARS-CoV-2 pandemic. *Swiss Med. Wkly.* **150**, (2020).
36. Lauring, A. S. & Hodcroft, E. B. Genetic Variants of SARS-CoV-2—What Do They Mean? *JAMA* (2021) https://doi.org/10.1001/jama.2020.27124.




**SeqCOVID-SPAIN consortium**

**Iñaki Comas[12,13], Fernando González-Candelas[13,15], Galo Adrian Goig[12], Álvaro Chiner-Oms[12], Irving Cancino-Muñoz[12], Mariana Gabriela López[12], Manuela Torres-Puente[12], Inmaculada Gomez-Navarro[12], Santiago Jiménez-Serrano[12], Lidia Ruiz-Roldán[15], María Alma Bracho[13,15], Neris García-González[15], Llúcia Martínez-Priego[16], Inmaculada Galán-Vendrell[16], Paula Ruiz-Hueso[16], Griselda De Marco[16], Maria Loreto Ferrús[16], Sandra Carbó-Ramírez[16], Giuseppe D'Auria[13,16], Mireia Coscollá[17], Paula Ruiz-Rodríguez[17], Francisco Javier Roig-Sena[18], Isabel Sanmartín[19], Daniel Garcia-Souto[20,21,22], Ana Pequeno-Valtierra[20], Jose M. C. Tubio[20,21], Jorge Rodríguez-Castro[20], Nuria Rabella[23,24,25], Ferrán Navarro[23,24,25], Elisenda Miró[23,24], Manuel Rodríguez-Iglesias[26,27,28], Fátima Galán-Sanchez[26,27,28], Salud Rodriguez-Pallares[26,27], María de Toro[29], María Bea Escudero[29], José Manuel Azcona-Gutiérrez[30], Miriam Blasco Alberdi[30], Alfredo Mayor[13,31,32,33], Alberto L. García-Basteiro[31,32,33], Gemma Moncunill[31,33], Carlota Dobaño[31,33], Pau Cisteró[31,33], Darío García-de-Viedma[34,35,36], Laura Pérez-Lago[34,35], Marta Herranz[34,35,36], Jon Sicilia[34,35], Pilar Catalán-Alonso[34,35,36], Patricia Muñoz[34,35,36], Cristina Muñoz-Cuevas[37,38], Guadalupe Rodríguez-Rodríguez[37,38], Juan Alberola-Enguidanos[39,40,41], Jose Miguel Nogueira[39,40,41], Juan José Camarena[39,40,41], Antonio Rezusta[42,43,44], Alexander Tristancho-Baró[42,43], Ana Milagro[42], Nieves Felisa Martínez-Cameo[42], Yolanda Gracia-Grataloup[42], Elisa Martró[13,45], Antoni E. Bordoy[45], Anna Not[45], Adrián Antuori-Torres[45], Rafael Benito[44,46], Sonia Algarate[44,46], Jessica Bueno[46], Jose Luis del Pozo[47], Jose Antonio Boga[48,49], Cristián Castelló-Abietar[48,49], Susana Rojo-Alba[48,49], Marta Elena Alvarez-Argüelles[48,49], Santiago Melon[48,49], Maitane Aranzamendi-Zaldumbide[50,51], Andrea Vergara-Gómez[52], Jovita Fernández-Pinero[53], Miguel J. Martínez[31,54], Jordi Vila[31,54], Elisa Rubio[31,54], Aida Peiró-Mestres[31,54], Jessica Navero-Castillejos[31,54], David Posada[55,56,57], Diana Valverde[55,56,57], Nuria Estévez-Gómez[55], Iria Fernandez-Silva[55,56], Loretta de Chiara[55,56], Pilar Gallego-García[55], Nair Varela[55], Rosario Moreno[58], Maria Dolores Tirado[58], Ulises Gomez-Pinedo[59], Mónica Gozalo-Margüello[60], Maria Eliecer-Cano[60], José Manuel Méndez-Legaza[60], Jesus Rodríguez-Lozano[60], María Siller[60], Daniel Pablo-Marcos[60], Antonio Oliver[61,62], Jordi Reina[61], Carla López-Causapé[61,62], Andrés Canut-Blasco[63], Silvia Hernáez-Crespo[63], Maria Luz A. Cordón[63], María-Concepción Lecároz-Agara[63], Carmen Gómez-González[63], Amaia Aguirre-Quiñonero[63], José Israel López-Mirones[63], Marina Fernández-Torres[63], Maria Rosario Almela-Ferrer[63], Nieves Gonzalo-Jiménez[64], Maria Montserrat Ruiz-García[64,65], Antonio Galiana[64,66], Judith Sanchez-Almendro[64,66], Gustavo Cilla[67], Milagrosa Montes[67], Luis Piñeiro[67], Ane Sorarrain[67], José María Marimón[67], Maria Dolores Gomez-Ruiz[68], José Luis López-Hontangas[68], Eva M. González Barberá[68], José María Navarro-Marí[69,70], Irene Pedrosa-Corral[69,70], Sara Sanbonmatsu-Gámez[69,70], Carmen Pérez-González[71], Francisco Chamizo-López[71], Ana Bordes-Benítez[71], David Navarro[41,72], Eliseo Albert[72], Ignacio Torres[72], Isabel Gascón[73], Cristina Juana Torregrosa-Hetland[73], Eva Pastor-Boix[73], Paloma Cascales-Ramos[73], Begoña Fuster-Escrivá[74], Concepción Gimeno-Cardona[41,74], María Dolores Ocete[74], Rafael Medina-Gonzalez[74], Julia González-Cantó[75], Olalla Martínez-Macias[75], Begoña Palop-Borrás[76], Inmaculada de Toro[76], Maria Concepción Mediavilla-Gradolph[76], Mercedes Pérez-Ruiz[76], Óscar González-Recio[77], Mónica Gutiérrez-Rivas[77], Encarnación Simarro-Córdoba[78], Julia Lozano-Serra[78], Lorena Robles-Fonseca[78], Adolfo de Salazar[79], Laura Viñuela-González[79], Natalia Chueca[79], Federico García[79], Cristina Gómez-Camarasa[79], Ana Carvajal[80], Raul de la Puente[80], Vicente Martín-Sánchez[13,81], Juan-Miguel Fregeneda-Grandes[80], Antonio José Molina[81], Héctor Argüello[80], Tania Fernández-Villa[81], Maria Amparo Farga-Martí[82], Victoria Domínguez-Márquez[82], José Javier Costa-Alcalde[83], Rocío Trastoy[83], Gema Barbeito-Castiñeiras[83], Amparo Coira[83], María Luisa Pérez-del-Molino[83], Antonio Aguilera[83], Anna M. Planas[84], Alex Soriano[85], Israel Fernandez-Cádenas[86], Jordi Pérez-Tur[12], Maria Ángeles Marcos[33,87], Antonio Moreno-Docón[88], Esther Viedma[89], Jesús Mingorance[90], Juan Carlos Galán-Montemayor[91] & Mónica Parra-Grande[92]**

[16]FISABIO, Servicio de Secuenciación, València, Spain. [17]Instituto de Biología Integrativa de Sistemas, I2SysBio (CSIC-Universitat de València), Valencia, Spain. [18]Servicio de Vigilancia y

Control Epidemiológico. Dirección General de Salud Pública y Adicciones. Conselleria de Sanitat Universal i Salut Pública. Generalitat Valenciana, Valencia, Spain. [19]Real Jardín Botánico, Consejo Superior de Investigaciones Científicas, Madrid, Spain. [20]Genomes and Disease, Centre for Research in Molecular Medicine and Chronic Diseases (CIMUS), Universidade de Santiago de Compostela, Santiago de Compostela, Spain. [21]Department of Zoology, Genetics and Physical Anthropology, Universidade de Santiago de Compostela, Santiago de Compostela, Spain. [22]Cancer Ageing and Somatic Mutation Programme, Wellcome Sanger Institute, Cambridge, CB1 8PS, UK. [23]Servei de Microbiologia.Hospital de la Santa Creu i Sant Pau, Barcelona, Spain. [24]CREPIMC. Institut d'Investigació Biomèdica Sant Pau, Barcelona, Spain. [25]Departament de Genètica i Microbiologia, Universitat Autònoma de Barcelona, Cerdanyola, Spain. [26]Servicio de Microbiología, H.U. Puerta del Mar, Cádiz, Spain. [27]INIBICA, Instituto de Investigación Biomédica de Cádiz, Cádiz, Spain. [28]Departamento de Biomedicina, Biotecnología y Salud Pública. Facultad de Medicina, Universidad de Cádiz, Cádiz, Spain. [29]Plataforma de Genómica y Bioinformática, Centro de Investigación Biomédica de La Rioja (CIBIR), Logroño, Spain. [30]Laboratorio de Microbiología. Hospital San Pedro, Logroño, Spain. [31]ISGlobal, Institute for Global Health, Barcelona, Spain. [32]Centro de Investigação em Saúde de Manhiça (CISM), Maputo, Mozambique. [33]Microbiology Department, Hospital Clinic I Provincial de Barcelona, Barcelona, Spain. [34]Servicio de Microbiología Clínica y Enfermedades Infecciosas. Hospital General Universitario Gregorio Marañón, Madrid, Spain. [35]Instituto de Investigación Sanitaria Gregorio Marañón, Madrid, Spain. [36]CIBER Enfermedades Respiratorias (CIBERES), Madrid, Spain. [37]Servicio de Microbiología Clínica. Hospital San Pedro de Alcántara, Cáceres, Spain. [38]Servicio Extremeño de Salud, Badajoz, Spain. [39]Servicio de Microbiología. Hospital Dr Peset, Valencia, Spain. [40]Conselleria de Sanitat i Consum. Generalitat Valenciana, Valencia, Spain. [41]Departamento Microbiología, Facultad de Medicina, Universidad de Valencia, Valencia, Spain. [42]Servicio de Microbiología Clínica Hospital Universitario Miguel Servet, Zaragoza, Spain. [43]Instituto de Investigación Sanitaria de Aragón, Centro de Investigación Biomédica de Aragón (CIBA), Zaragoza, Spain. [44]Facultad de Medicina, Universidad de Zaragoza, Zaragoza, Spain. [45]Servicio de Microbiología, Laboratori Clínic Metropolitana Nord, Hospital Universitari Germans Trias i Pujol, Badalona, Barcelona, Spain. [46]Hospital Clínico Universitario Lozano Blesa, Zaragoza, Spain. [47]Servicio de Enfermedades Infecciosas y Microbiología clínica. Clínica Universidad de Navarra, Pamplona, Spain. [48]Servicio de Microbiología, Hospital Universitario Central de Asturias, Oviedo, Spain. [49]Grupo de Microbiología Traslacional Instituto de Investigación Sanitaria del Principado de Asturias (ISPA), Oviedo, Spain. [50]Servicio de Microbiología, Hospital Universitario Cruces, Bilbao, Spain. [51]Grupo de Microbiología y Control de Infección Instituto de Investigación Sanitaria Biocruces Bizkaia, Bizkaia, Spain. [52]Servicio de Microbiología & CORE de Biología Molecular, CDB, Hospital Clínic, Barcelona, Spain. [53]Centro de Investigación en Sanidad Animal. Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria, O.A., M.P. - INIA, Valdeolmos, Spain. [54]Departamento de Microbiología, Hospital Clínic de Barcelona, Barcelona, Spain. [55]CINBIO, Universidade de Vigo, Vigo, Spain. [56]Department of Biochemistry, Genetics, and Immunology, Universidade de Vigo, Vigo, Spain. [57]Galicia Sur Health Research Institute (IIS Galicia Sur), SERGAS-UVIGO, Vigo, Spain. [58]Hospital General Universitario de Castellón, Castellón, Spain. [59]IdISSC/Hospital Clínico San Carlos, Madrid, Spain. [60]Hospital Marqués de Valdecilla - IDIVAL, Santander, Spain. [61]Servicio de Microbiología, Hospital Universitario Son Espases, Palma de Mallorca, Spain. [62]Instituto de Investigación Sanitaria de las Islas Baleares, Baleares, Spain. [63]Servicio de Microbiología, Hospital Universitario de Álava, Osakidetza-Servicio Vasco de Salud, Vitoria-Gasteiz (Álava), Spain. [64]Servicio Microbiología, Departamento de Salud de Elche-Hospital General, Elche, Alicante, Spain. [65]Departamento de Producción Vegetal y Microbiología, Universidad Miguel Hernández, Elche, Spain. [66]Fundación para el Fomento de la Investigación Sanitaria y Biomédica de la Comunitat Valenciana:, Elche, Alicante, Spain. [67]Biodonostia; Osakidetza, Hospital Universitario Donostia, Servicio de Microbiología, San Sebastián, Spain. [68]Hospital Universitario y Politécnico La Fe, Servicio de Microbiología, Valencia, Spain. [69]Servicio de Microbiología, Hospital Universitario Virgen de las Nieves, Granada, Spain. [70]Hospital Universitario Virgen de las Nieves, Instituto de Investigación Biosanitaria ibs, Granada, Spain. [71]Hospital Universitario de Gran Canaria Dr. Negrin, Las Palmas de Gran Canaria, Spain. [72]Microbiology Service, Hospital Clínico Universitario, INCLIVA Research Institute, Valencia, Spain. [73]Laboratorio de Microbiología, Hospital General Universitario de Elda, Elda, Alicante, Spain. [74]Servicio de Microbiología, Consorcio Hospital General Universitario de Valencia, Valencia, Spain. [75]Laboratorio Biología Molecular, Área de Diagnóstico Biológico, Hospital Universitario La Ribera, Alzira, Valencia, Spain. [76]Servicio de Microbiologia, Hospital Regional Universitario de Málaga, Málaga, Spain. [77]Instituto Nacional de Investigación y Tecnología Agraria y Alimentaria, O.A., M.P. - INIA, Madrid, Spain. [78]Hospital General Universitario de Albacete, Albacete, Spain. [79]Hospital Universitario San Cecilio, Granada, Spain. [80]Animal Health Department, Universidad de León, León, Spain. [81]Research Group on Gene-Environment Interactions and Health. Institute of Biomedicine (IBIOMED). Universidad de León, León, Spain. [82]Servicio de Microbiología, Hospital Arnau de Vilanova, Valencia, Spain. [83]Hospital Clínico Universitario de Santiago de Compostela, Santiago de Compostela, Spain. [84]Biomedical Research Institute of Barcelona (IIBB), Spanish National Research Council (CSIC), Barcelona, Spain. [85]Servicio de Enfermedades Infecciosas, Hospital Clínic de Barcelona, Barcelona, Spain. [86]Biomedical Research Institute Sant Pau (IIB Sant Pau), Barcelona, Spain. [87]Institut of Global Health of Barcelona (ISGlobal), Barcelona, Spain. [88]Servicio de Microbiología, Hospital Clínico Universitario Virgen de la Arrixaca, Departamento de Genética y Microbiología, Universidad de Murcia, Carretera Madrid-Cartagena sn, 30120- El Palmar, Murcia, Spain. [89]Hospital Universitario 12 de Octubre, Madrid, Spain. [90]Hospital Universitario La Paz, Madrid, Spain. [91]Hospital Universitario Ramón y Cajal, Madrid, Spain. [92]Laboratorio de Microbiología, Hospital Marina Baixa, Villajoyosa, Spain.

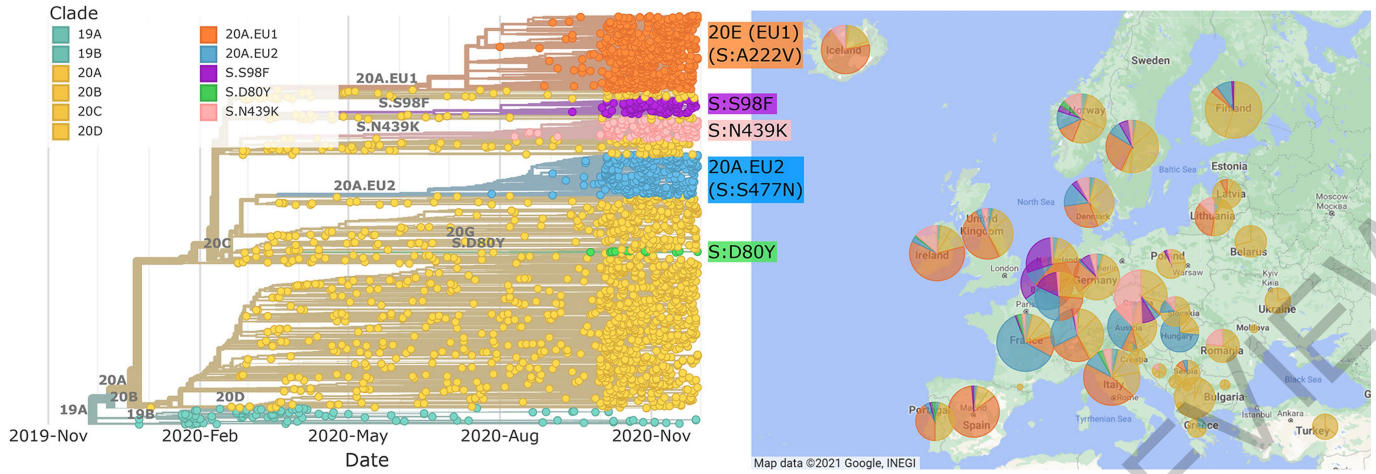



**Fig. 1 | Phylogenetic overview of SARS-CoV-2 in Europe through the end of November.** The tree shows a representative sample of isolates from Europe colored by clade and by the variants highlighted in this paper. Clade 20A and its daughter clades 20B and 20C have variant S:D614G and are colored in yellow. A novel variant (orange; 20E (EU1)) with mutation S:A222V on a S:D614G background emerged in early summer and is common in most countries with recent sequences. A separate variant (20A.EU2, blue) with mutation S:S477N is prevalent in France. On the right, the proportion of sequences belonging to each variant (through the end of November) is shown per country. Tree and visualization were generated using the Nextstrain platform (Hadfield et al., 2018) as described in methods.

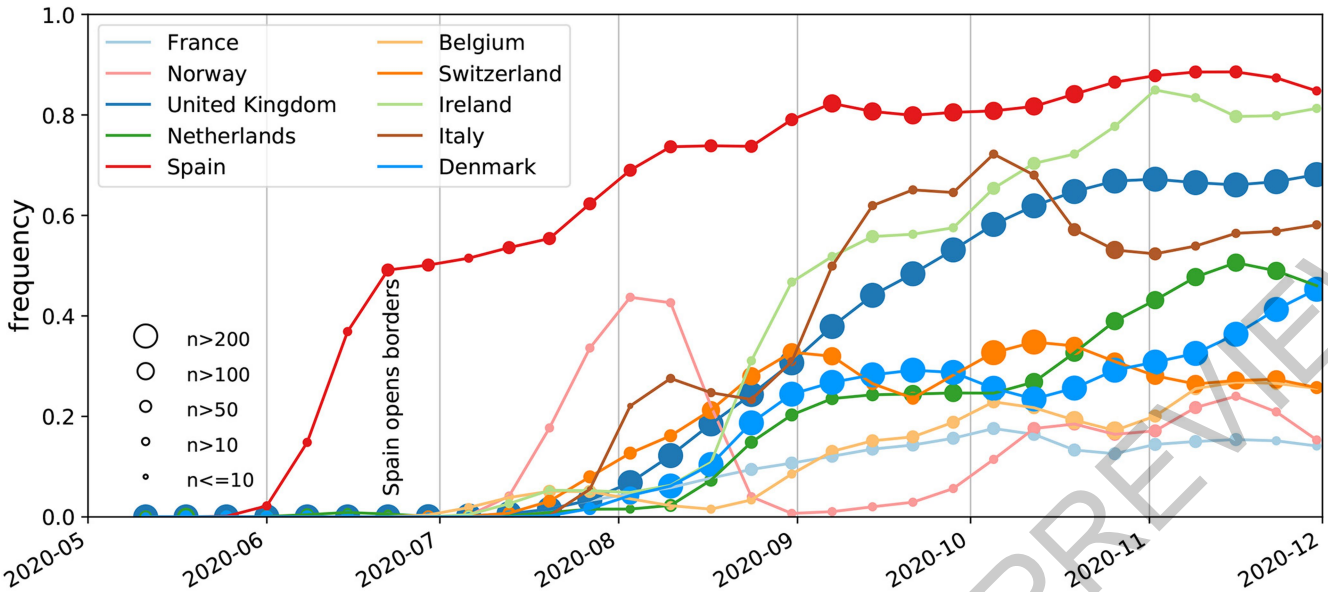



**Fig. 2 | Frequency of submitted samples that are 20E (EU1) in selected countries.** We include the eight countries which have at least 200 sequences from 20E (EU1), as well as Norway and France, to illustrate points in the text. The symbol size indicates the number of available sequence by country and time point in a non-linear manner. In most countries we observe a gradual rise from mid-July settling to a plateau. In contrast, Norway observed a sharp peak in summer but seems to have brought cases down quickly, though they began growing again in September. When the last data point included only very few sequences, it has been dropped for clarity. Frequencies are smoothing using a Gaussian with $\sigma = 1w$.

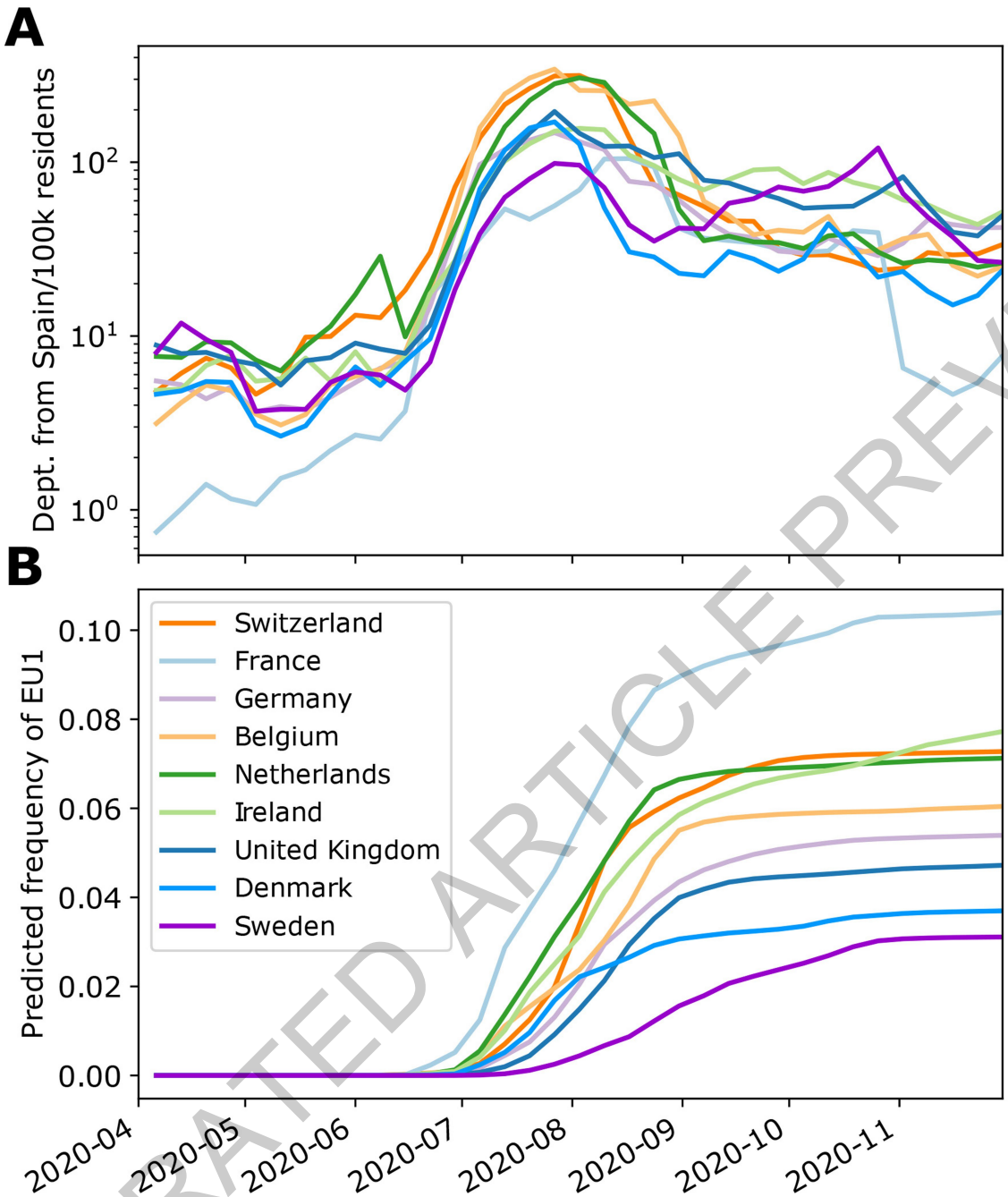



**Fig. 3 | Travel volume and contribution of imported infections.** Travel from Spain to other European countries resumed in July (though low compared to previous years). Assuming that travel returnees are infected at the average incidence of Spanish province they visited and transmit the virus at the rate of their resident population, imports from Spain are expected to account between 2 and 12% of SARS-CoV-2 cases after the summer. Traveler incidence is calculated using case and travel data at the level of provinces. Note that this model only accounts for contribution of summer travel and that stochastic fluctuations and other variants after the summer will results in further variation in the frequency of 20E (EU1). See Methods and Fig. S8.

# Methods

### Phylogenetic analysis

We use the Nextstrain pipeline for our phylogenetic analyses https://github.com/ nextstrain/ncov/ [4]. Briefly, we align sequences using mafft[37], subsample sequences (see below), add sequences from the rest of the world for phylogenetic context based on genomic proximity, reconstruct a phylogeny using IQTree[38] and infer a time scaled phylogeny using TreeTime[39]. For computational feasibility, ease of interpretation, and to balance disparate sampling efforts between countries, the Nextstrain-maintained runs sub-sample the available genomes across time and geography, resulting in final builds of ~5,000 genomes each. After sub-sampling, the 20E (EU1) cluster within the Nextstrain build contains 5,145 sequences, 3,369 of which are unique (accounting for Ns).

Sequences were downloaded from GISAID at the end of January and analyzed using the nextstrain/ncov workflow, using a cutoff date of the 30 Sept (Fig S4a) or 30 Nov (all other analyses). These dates were chosen to focus first on the introductions over the summer (for 30 Sept) and then to highlight ongoing circulation through the autumn (30 Nov) prior to the spread of the variants of concern identified in December 2020 and January 2021. A table acknowledging the invaluable contributions by many labs is available as a supplement. The Swiss SARS-CoV-2 sequencing efforts are described in Nadeau et al.[40] and Stange et al.[41]. The majority of Swiss sequences used here are from the Nadeau et al.[40] data set, the remainder are available on GISAID.

### Defining the 20E (EU1) Cluster

The cluster was initially identified as a monophyletic group of sequences stemming from the larger 20A clade with amino acid substitutions at positions S:A222V, ORF10:V30L, and N:A220V or ORF14:L67F (overlapping reading frame with N), corresponding to nucleotide mutations C22227T, C28932T, and G29645T. In addition, sequences in 20E (EU1) differ from their ancestors by the synonymous mutations T445C, C6286T, and C26801G.

The sub-sampling of the standard Nextstrain analysis means that we are not able to visualize the true size or phylogenetic structure of the cluster in question. To specifically analyze this cluster using almost all available sequences, we designed a specialized build which focuses on cluster-associated sequences and their most genetically similar neighbors. For computational reasons, we limit the number of samples to 900 per country per month. As only the UK has more sequences than this in the relevant time period, this results in a random downsampling of sequences from the UK for the months of August, September, and October. Further, we excluded several problematic sequences due to high intra-sample variation, wrong dates, and over-divergence (divergence values are implausible given the provided dates). A full list of the sequences excluded (and the reason why) is given on github in "bad_sequences.py."

We identify sequences in the cluster based on the presence of nucleotide substitutions at positions 22227, 28932, and 29645 and use this set as a 'focal' sample in the nextstrain/ncov pipeline. This selection will exclude any sequences with no coverage or reversions at these positions, but the similarity-based sampling during the Nextstrain run will identify these, as well as any other nearby sequences, and incorporate them into the dataset. We used these three mutations as they included the largest number of sequences that are distinct to the cluster. By this criterion, there are currently 60,316 sequences in the cluster sampled before 30 November 2020.

To visualize the changing prevalence of the cluster over time, we plotted the proportion of sequences identified by the four substitutions described above as a fraction of the total number of sequences submitted, per ISO week. Frequencies of other clusters are identified in an analogous way.

### Phylogeny and Geographic Distribution

The size of the cluster and number of unique mutations among individual sequences means that interpreting overall patterns and connections between countries is not straightforward. We aimed to create a simplified version of the tree that focuses on connections between countries and de-emphasizes onward transmissions within a country. As our focal build contains 'background' sequences that do not fall within the cluster, we used only the monophyletic clade containing the four amino-acid changes and three synonymous nucleotide changes that identify the cluster. Then, subtrees that only contain sequences from one country were collapsed into the parent node. The resulting phylogeny contains only mixed-country nodes and single-country nodes that have mixed-country nodes as children. (An illustrative example of this collapsing can be seen in Fig. ED4(b-d).) Nodes in this tree thus represent ancestral genotypes of subtrees: sequences represented within a node may have further diversified within their country, but share a set of common mutations. We count all sequences in the subtrees towards the geographic distribution represented in the pie-charts in Fig. ED4a.

This tree allows us to infer lower bounds for the number of introductions to each country, and to identify plausible origins of those introductions. It is important to remember that, particularly for countries other than the UK, the full circulating diversity of the variant is probably not being captured, thus intermediate transmissions cannot be ruled out. In particular, the closest relative of a particular sequence will often have been sampled in the UK simply because sequencing efforts in the UK exceed most other countries by orders of magnitude. It is, however, not our goal to identify all introductions but to investigate large scale patterns of spread in Europe.

### Travel volume and destination

Mobile phone roaming data were used to estimate the number of visitors from a given country departing from a given province for each calendar week. The mobile phone record data set contains approximately 13 million devices, with over 2.6 million roamers. A visitor is considered to be departing the country during a given week if they are not seen in the data set for the next eight weeks. The nationality of a visitor is inferred from the Mobile Country Code (MCC). The total number of unique visitors is aggregated for each province and each week in the period of study; these totals are then scaled using official statistics as reference to account for the partial coverage of data set.

### Estimation of contributions from imports

To estimate how the frequency of 20E (EU1) is expected to change in country $X$ due to travel, we consider the following simple model: A fraction $\alpha_i$ of the population of $X$ returns from Spain every week $i$ (estimated from roaming data, see above) and is infected with 20E (EU1) with a probability $p_i$ given by its per capita weekly incidence in Spain. Incidence is the weighted average over incidence in Spanish provinces by the distribution of visitors across the provinces. The week-over-week fold change of the epidemic within $X$ is calculated as $g_i = (c_i - \alpha_i p_i)/c_{i-1}$, where $c_i$ is the per capita incidence in week $i$ in $X$. This fold-change captures the local growth of the epidemic in country $X$. The total number of 20E (EU1) cases $v_i$ in week $i$ is hence $v_i = g_i v_{i-1} + p_i \alpha_i$, while the total number of non-20E (EU1) cases is $r_i = g_i r_{i-1}$. Running this recursion from mid-June to November results in the frequency trajectories in Fig. 3.

From 1 June 2020 to 30 September 2020, the Swiss Federal Office of Public Health (FOPH) reported 23,199 confirmed SARS-CoV-2 cases. 14,583 (62.9%) cases provided information about their likely place of exposure and country of infection in a clinical registration form. Of these, 3,304 (22.7%) reported an exposure abroad and 136 (0.9%) named Spain as the country of infection. The Robert-Koch-Institute reported statistics on likely country of infection by calendar week in their daily situation reports[42].

### Phylodynamic analysis of Swiss transmission chains

We identified introductions into Switzerland and downstream Swiss transmission chains by considering a tree of all available Swiss sequences combined with foreign sequences with high similarity to Swiss sequences (full procedure described in Nadeau et al. (2020)[40]. Putative transmission chains were defined as majority Swiss clades allowing for at most 3 "exports" to third countries. Identification of transmission chains is complicated by polytomies in SARS-CoV-2 phylogenies and we bounded the resulting uncertainty by either (i) considering all subtrees descending from the polytomy as separate introductions (called 'max' in Fig ED8) and (ii) aggregating all into a single introduction (called 'min'), see Nadeau et al. (2020) for details. We further extended this analysis to include a pan-European dataset consisting of putative transmission chains defined via the collapsed phylogenies discussed earlier in the methods. Specifically, each section of a pie graph, which corresponds to a country-specific collection of sequences, was taken as a single introduction. Non-20E (EU1) $R_e$ estimates were obtained from case data and the estimated frequency of 20E (EU1) in different countries.

The phylodynamic analysis of the transmission chains was performed using BEAST2 with a birth-death-model tree prior[43,44]. 20E (EU1) and non-20E (EU1) variants share a sampling probability and log$R_e$ has an Ornstein-Uhlenbeck prior, see Nadeau et al. (2020)[40] for details (but note a different smoothing prior is used there).

### Enzyme-linked immunosorbent assay (ELISA)

384-well Maxisorp plates (Thermo Fisher) were coated overnight at room temperature with 3 $\mu$g/mL in 20mM Tris pH 8 and 150mM NaCl of SARS-CoV-2 S2P[45] or SARS-CoV-2 A222V-D614G S2P, produced as previously described in Walls et al. (2020). Briefly, Expi293F cells were transiently transcribed with a plasmid containing the spike protein and supernatant was clarified six days later prior to Ni Sepharose resin purification and flash freezing. Gibco (Fisher) Expi293F Cells were used for protein production and have not been authenticated or tested for mycoplasma contamination. They are not in the database of commonly misidentified cell lines. Plates were slapped dry and blocked with Blocker Casein in TBS (Thermo Fisher) for one hour at 37 °C. Plates were slapped dry and S2E12[18] or S309[22] antibodies were serially diluted 1:3 with a starting concentration of 1000nM in TBST or NIBSC human plasma (20/130 https://www.nibsc.org/documents/ifu/20-130.pdf) was serially diluted 1:3 starting at 1:4 of original concentration in TBST and added to the plate for one hour at 37 °C. Plates were washed 4x with TBST using a 405 TS Microplate Washer (BioTek) followed by addition of 1:5,000 goat anti-human Fc IgG-HRP (Thermo Fisher) for one hour at 37 °C. Plates were washed 4x and TMB Microwell Peroxidase (Seracare) was added. The reaction was quenched after 1-2 minutes with 1 N HCl and the A450 of each well was read using a Varioskan Lux plate reader (Thermo Fisher).

### Pseudotyped Lentivirus Production and Titering

The S:A222V mutation was introduced into the protein-expression plasmid HDM-Spiked21-D614G, which encodes a codon-optimized spike from Wuhan-Hu-1 (Genbank NC 045512) with a 21-amino acid cytoplasmic tail deletion and the D614G mutation (Greaney et al., 2020). This plasmid is also available on AddGene (plasmid 158762). We made two different versions of the A222V mutant that differed only in which codon was used to introduce the valine mutation (either GTT or GTC). The sequences of these plasmids (HDM Spike-d21D614G-A222V-GTT and HDM Spike-d21-D614G-A222V-GTC) are available as supplement files at github.com/emmahodcroft/cluster_scripts/plasma_data.

Spike-pseudotyped lentiviruses were produced as described in[46]. Two separate plasmid preps of the A222V (GTT) spike and one plasmid prep of the A222V (GTC) spike were each used in duplicate to produce six replicates of A222V spike-pseudotyped lentiviruses. Three plasmid preps of the initial D614G spike plasmid (with the 21-amino acid cytoplasmic tail truncation) were each used once used to make three replicates of D614G spike-pseudotyped lentiviruses. All viruses were titered in duplicate.

Lentiviruses were produced with both Luciferase IRES ZsGreen and ZsGreen only lentiviral backbones[46], and then titered using luciferase signal or percentage of fluorescent cells, respectively. All viruses were titered in 293T-ACE2 cells (BEI NR-52511) as described in[46], with the following modifications. Viruses containing luciferase were titered starting at a 1:10 dilution followed by 5 serial 2-fold dilutions. The Promega BrightGlo luciferase system was used to measure relative luciferase units (RLUs) ~65 hours post-infection and RLUs per mL were calculated at each dilution then averaged across all dilutions for each virus. Viruses containing only ZsGreen were titered starting at a 1:3 dilution followed by 4 serial 5-fold dilutions. The 1:375 dilution was visually determined to be ~1% positive about 65 hours post-infection and was used to calculate the percent of infected cells using flow cytometry (BD FACSCelesta cell analyzer). Viral titers were then calculated using the percentage of green cells via the Poisson formula. To normalize viral titers by lentiviral particle production, p24 concentration (in pg/mL) was quantified by ELISA according to kit instructions (Advanced Bioscience Laboratories Cat. #5421). All viral supernatants were measured in technical duplicate at a 1:100,000 dilution.

### Reporting summary

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

## Data availability

Code used for the above analyses is available at github.com/neherlab/2020_EU1_paper. The code used to run the cluster builds is available at github.com/emmahodcroft/ncov_cluster. Sequence data were obtained from GISAID and tables listing all accession numbers of sequences are available as supplementary information.


37. Katoh, K., Misawa, K., Kuma, K. & Miyata, T. MAFFT: a novel method for rapid multiple sequence alignment based on fast Fourier transform. *Nucleic Acids Res*. **30**, 3059–3066 (2002).
38. 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 (2020).
39. Sagulenko, P., Puller, V. & Neher, R. A. TreeTime: Maximum-likelihood phylodynamic analysis. *Virus Evol*. **4**, (2018).
40. Nadeau, S. *et al*. Quantifying SARS-CoV-2 spread in Switzerland based on genomic sequencing data. *medRxiv* (2020) https://doi.org/10.1101/2020.10.14.20212621.
41. Stange, M. *et al*. SARS-CoV-2 phylogeny during the early outbreak in the Basel area, Switzerland: import and spread dominated by a single B.1 lineage variant (C15324T). *medRxiv* 2020.09.01.20186155 (2020) https://doi.org/10.1101/2020.09.01.20186155.
42. Robert-Koch-Institute. *RKI - Coronavirus SARS-CoV-2 - Aktueller Lage-/Situationsbericht des RKI zu COVID-19*. (2020).
43. Stadler, T., Kühnert, D., Bonhoeffer, S. & Drummond, A. J. Birth-death skyline plot reveals temporal changes of epidemic spread in HIV and hepatitis C virus (HCV). *Proc Natl Acad Sci U A* **110**, 228–233 (2013).
44. Bouckaert, R. *et al*. BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. *PLOS Comput. Biol*. **15**, e1006650 (2019).
45. Pallesen, J. *et al*. Immunogenicity and structures of a rationally designed prefusion MERS-CoV spike antigen. *Proc. Natl. Acad. Sci*. **114**, E7348–E7357 (2017).
46. Crawford, K. H. D. *et al*. Protocol and Reagents for Pseudotyping Lentiviral Particles with SARS-CoV-2 Spike Protein for Neutralization Assays. *Viruses* **12**, 513 (2020).



**Acknowledgements** We are grateful to researchers, clinicians, and public health authorities for making SARS-CoV-2 sequence data available in a timely manner. We also wish to thank the COVID-19 Genomics UK consortium for their notable sequencing efforts, which have provided a third of the sequences currently publicly available. We would also like to thank the Swiss Federal Office of Public Health (FOPH) for providing access to their data. This work was supported by the Swiss National Science Foundation (SNSF) through grant numbers 31CA30 196046 (to RAN, EBH, CLA), 31CA30 196267 (to TS), European Union's Horizon 2020 research and innovation programme - project EpiPose (No 101003688) (MLR, CLA), core funding by the University of Basel and ETH Zürich, the National Institute of General Medical Sciences (R01GM120553 to DV), the National Institute of Allergy and Infectious Diseases (DP1AI158186 and HHSN272201700059C to DV), a Pew Biomedical Scholars Award (DV), an Investigators in the Pathogenesis of Infectious Disease Awards from the Burroughs Wellcome Fund (DV and JDB), a Fast Grants (DV), and NIAID

grants R01AI141707 (JDB) and F30AI149928 (KHDC). SeqCOVID-SPAIN is funded by the Instituto de Salud Carlos III project COV20/00140, Spanish National Research Council and ERC StG 638553 to IC and BFU2017-89594R from MICIN to FGC. JDB is an Investigator of the Howard Hughes Medical Institute.

**Author contributions** EBH identified the cluster, led the analysis, created figures, and drafted the manuscript. RAN analyzed data, created figures, and drafted the manuscript. MZ, SN, TGV, CLA, TS, and MLR analyzed data and created figures. VD investigated structural aspects and created figures. JDB, JEB, ACW, DC, and KHDC performed experimental assays and created figures. IC and FGC interpreted the origin of the cluster and contributed data. DM and AH contributed and interpreted data. All authors contributed to and approved the final manuscript.

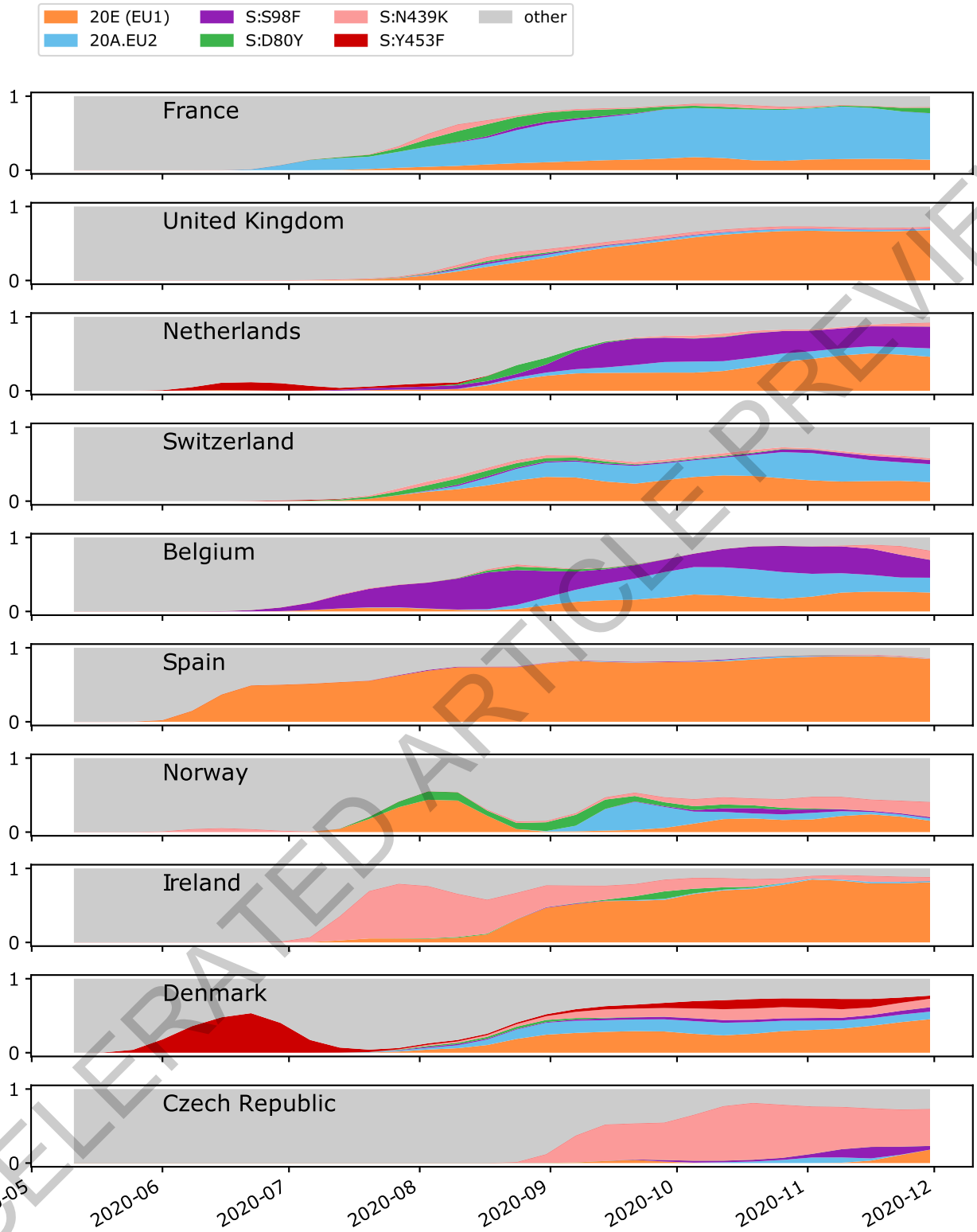
**Competing interests** DV is a consultant for Vir Biotechnology Inc. DC is an employee of Vir Biotechnology and may hold shares in Vir Biotechnology. The Veesler laboratory has received an unrelated sponsored research agreement from Vir Biotechnology Inc. AH is a co-founder of Kido Dynamics, DM is employed by Kido Dynamics. The other authors declare no competing interests.

**Additional information**

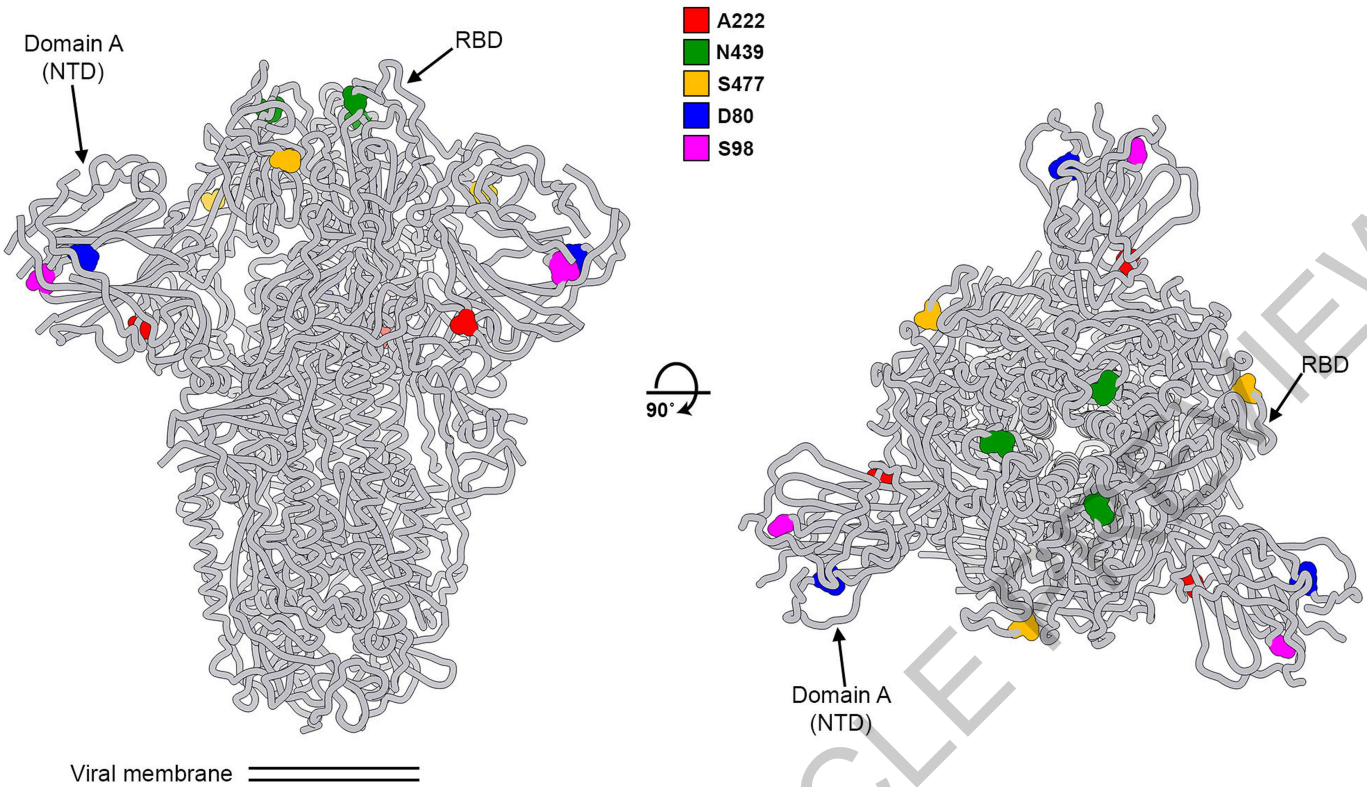
**Supplementary information** The online version contains supplementary material available at https://doi.org/10.1038/s41586-021-03677-y.

**Correspondence and requests for materials** should be addressed to E.B.H. or R.A.N.

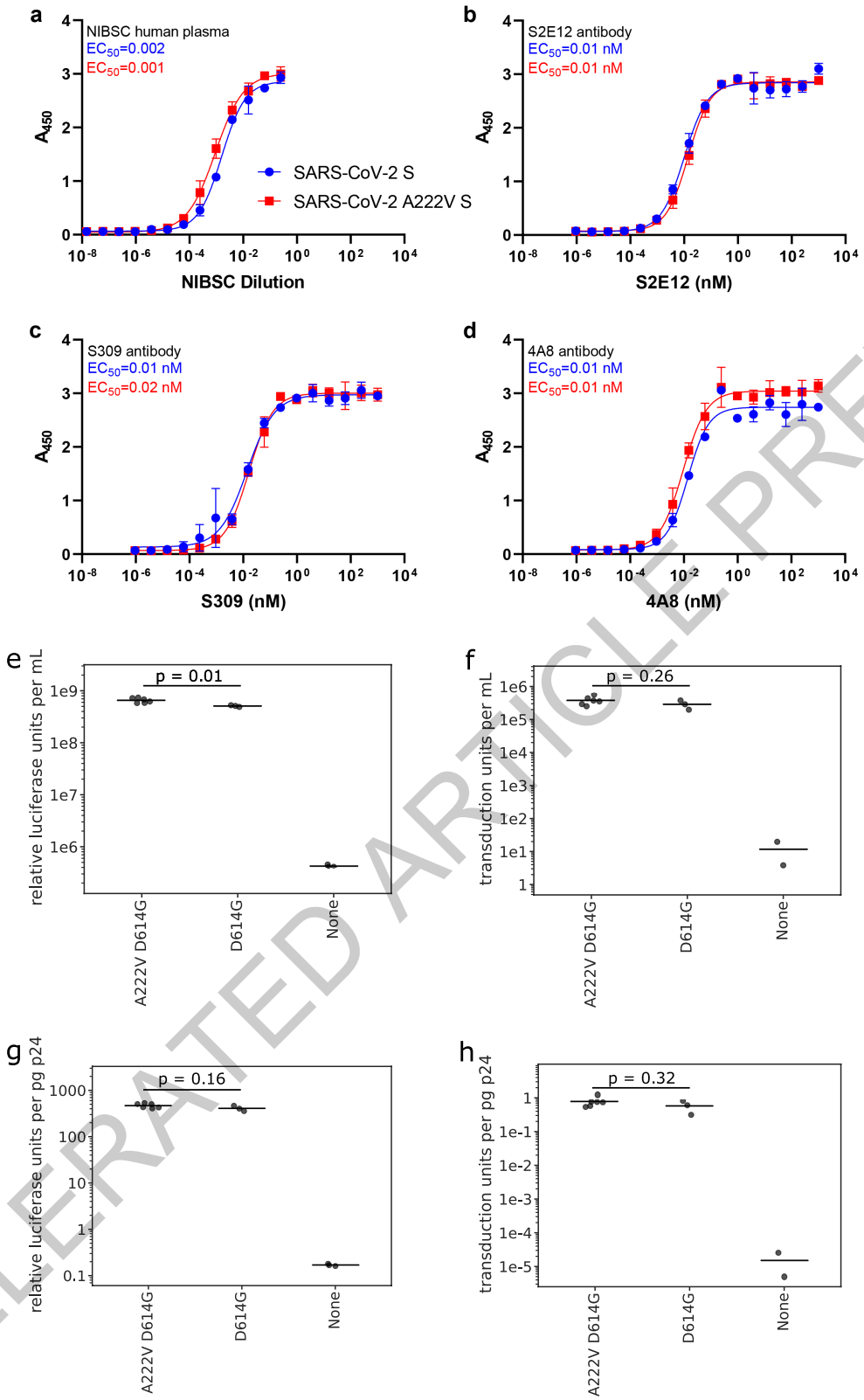
**Peer review information** *Nature* thanks the anonymous reviewers for their contribution to the peer review of this work. Peer reviewer reports are available.




**Extended Data Fig. 1 | Variant dynamics in different European Countries.** In countries with at least ten sequences that fall into any of the defined clusters, the proportion of sequences per ISO week that fall into each cluster is shown.

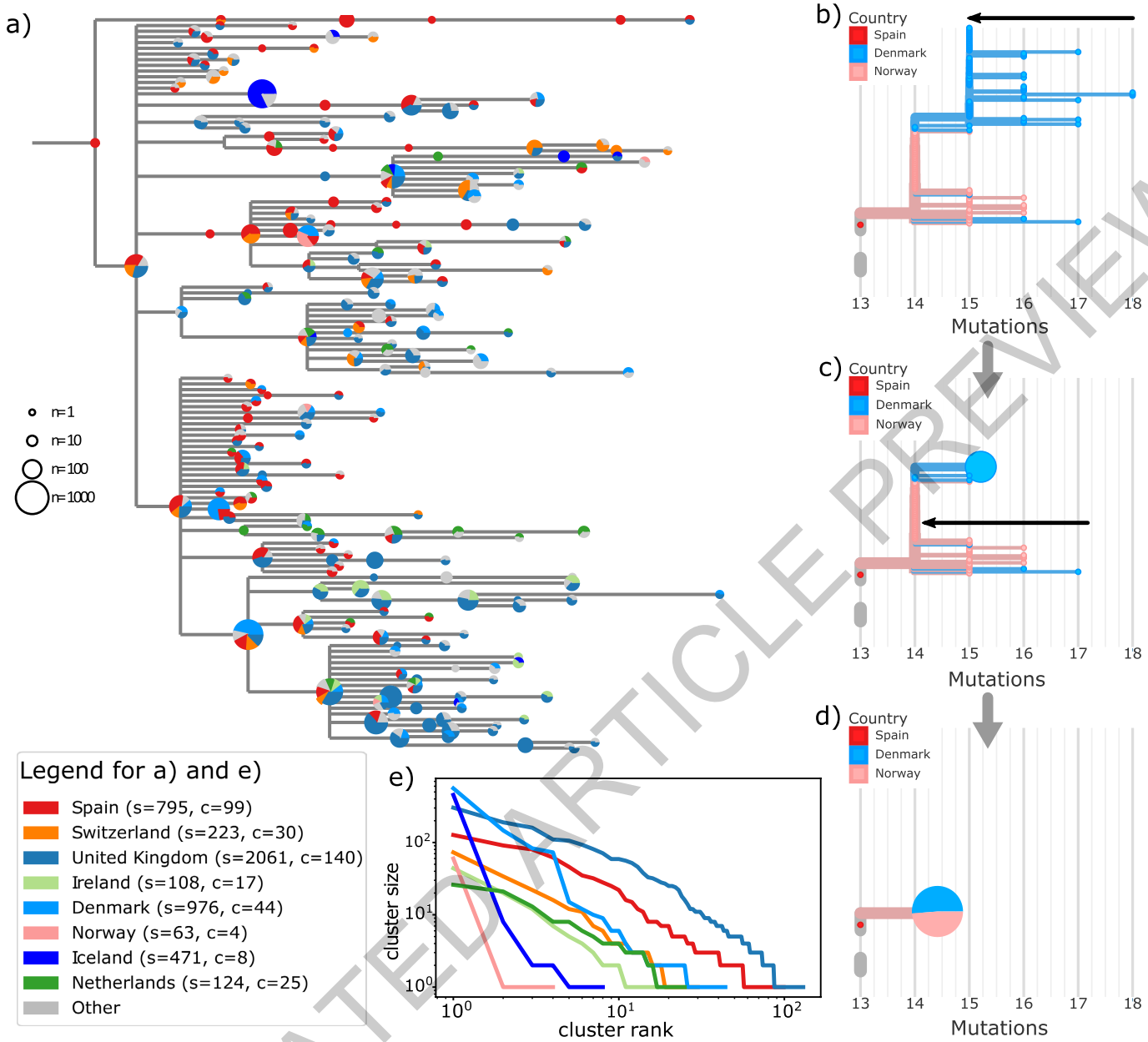



**Extended Data Fig. 2 | Structure model of the SARS-CoV-2 Spike Protein.** Two orthogonal orientations of the SARS-CoV-2 spike glycoprotein trimer highlighting the position of the variants described in the manuscript and the receptor binding domain (RBD) and the NTD (domain A). 222: red; 439: green; 477: orange; 80: blue; 98: magenta.

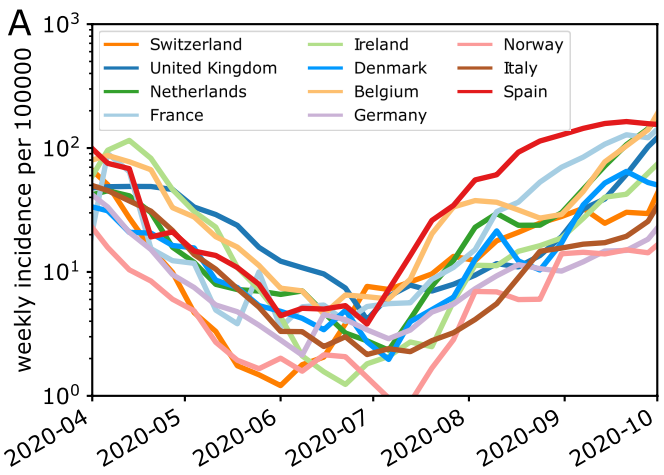


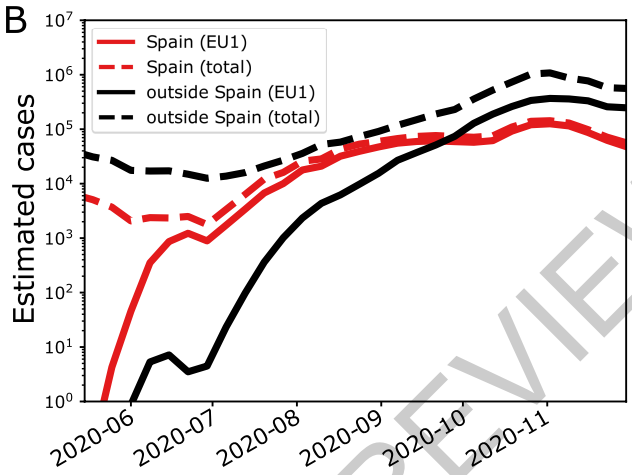

**Extended Data Fig. 3** | See next page for caption.

**Extended Data Fig. 3 | The substitution A222V in spike has no substantial effect on antigenic properties (a-d) and replication of pseudotyped lentiviruses (e-h).** (**a**) Binding of a serial dilution of NIBSC convalescent plasma to immobilized SARSCoV-2 2P S (blue) or SARS-CoV-2 2P A222V D614G S (red). (**b-c**), Binding of serially diluted concentrations of the human neutralizing antibodies S309 (**b**) and S2E12 (**c**) to immobilized SARSCoV-2 2P S (blue) or SARS-CoV-2 2P A222V D614G S (red). (**d**) Binding of serially diluted concentrations of the human neutralizing antibody 4A8 to immobilized SARS-CoV-2 2P S (blue) or SARS-CoV-2 2P A222V D614G S (red). $n = 2$ experiments performed with independent protein preparations (each in duplicate). Each data point consists of a technical duplicate of each antibody or plasma dilution, and the error bars show standard deviations. The experiment shown is representative of two independent experiments. (**e**) Titers of lentiviral particles carrying luciferase in the viral genome. The horizontal line indicates the mean titer. (**f**) Titers of lentiviral particles carrying the fluorescent protein ZsGreen in the viral genome. The horizontal line indicates the mean titer. In both cases, titers with the A222V mutation are on average higher by a factor 1.3. (**g**) Titers of lentiviral particles carrying luciferase in the viral genome normalized by the p24 concentration (pg/mL) of each viral supernatant. After p24 normalization, the titer difference shrinks fom 1.28 to 1.14 fold, increasing the p-value to 0.16. (**h**) Titers of lentiviral particles carrying ZsGreen in the viral genome normalized by the p24 concentration (pg/mL) of each viral supernatant. All *p*-values calculated using a two-sided *t*-test.

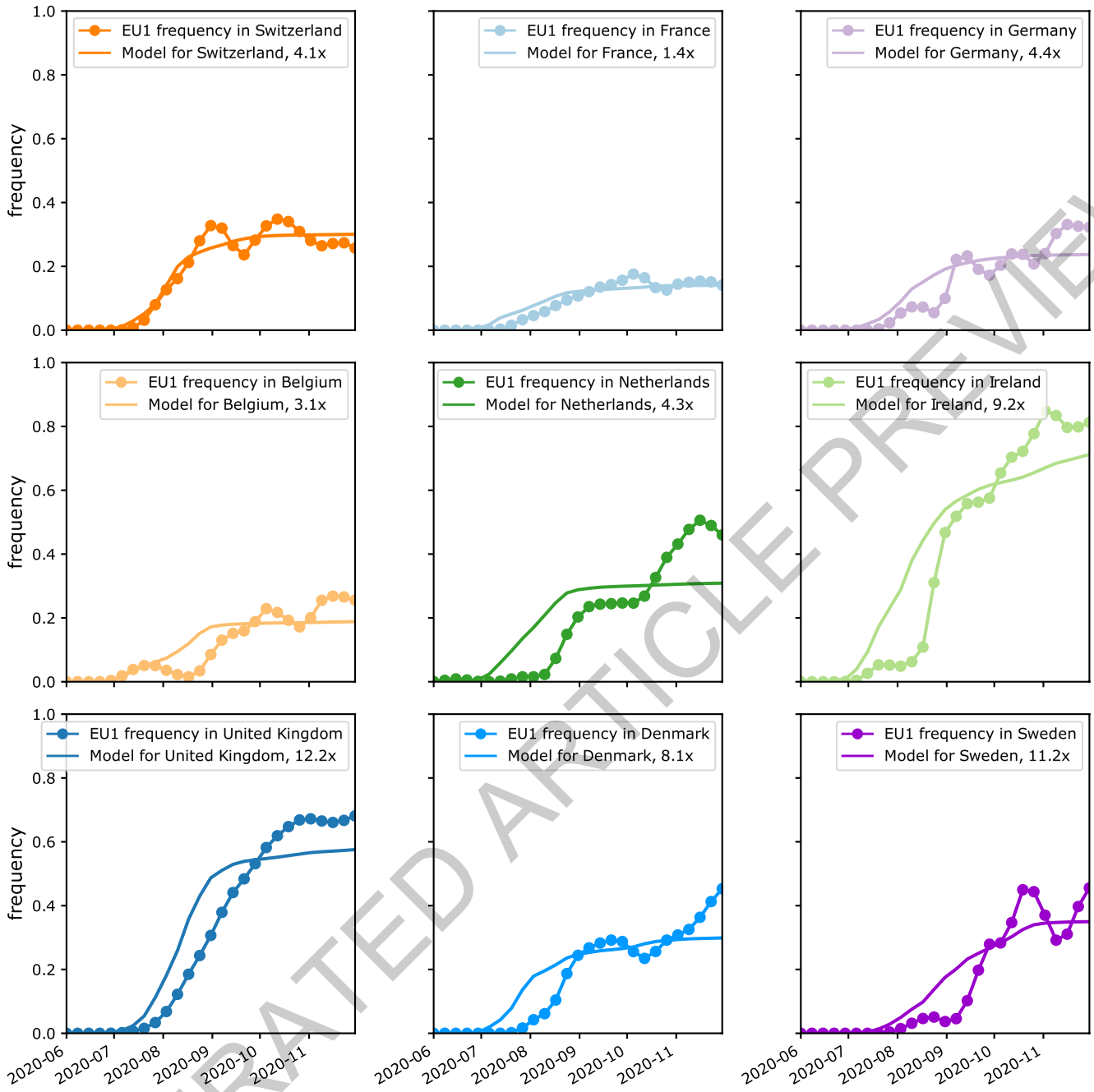



**Extended Data Fig. 4 | Collapsed genotype phylogeny and statistics of putative introductions.** (**a**) The phylogeny shown is the subtree of the 20E (EU1) cluster using data from samples collected before 30 Sept 2020 and available on GISAID as of Jan 2021, with sequences carrying all six defining mutations. Pie charts show the representation of sequences from selected countries at each node. Size of the pie chart indicates the total number of sequences at each node. Pie chart fractions scale non-linearly with the true counts (fourth root) to ensure all countries are visible and branch lengths are jittered to reduce overlap. Though the jitter means branch lengths should be interpreted with caution, the smallest branches shown in the tree equal to 1 mutation. See also Extended Data Figure 6 for an example of how collapsing was done. (**b-d**) Show an example of how the pie-chart phylogeny was created. The tree is shown in 'divergence view' with the branch lengths in mutations. Internal nodes are shown as horizontal lines with other nodes (internal and external) branching from them. If sequences are identical, they align on the horizontal line. In this example zooming in to the Norwegian cluster, the outermost tips are first collapsed down to their parental node (**b**), forming a pie chart that consists only of sequences from Denmark (**c**). This single-country pie chart is collapsed with the next level of nodes (**d**), including more sequences from Denmark and sequences from Norway, to form a multi-country pie chart. (**e**) Rank-order plots of sizes of clusters of sequences in the pie-chart slices, in different countries, compatible with a single introduction. Countries like Norway and Iceland have relatively few clusters, with one or two large clusters dominating, suggesting a small number of introductions dominated 20E (EU1) circulation. Countries like the UK and Denmark, on the other hand, show many clusters of varying size, indicating multiple introductions that led to onward spread. The legend indicates total number of sequences *s* and number of clusters *c*.

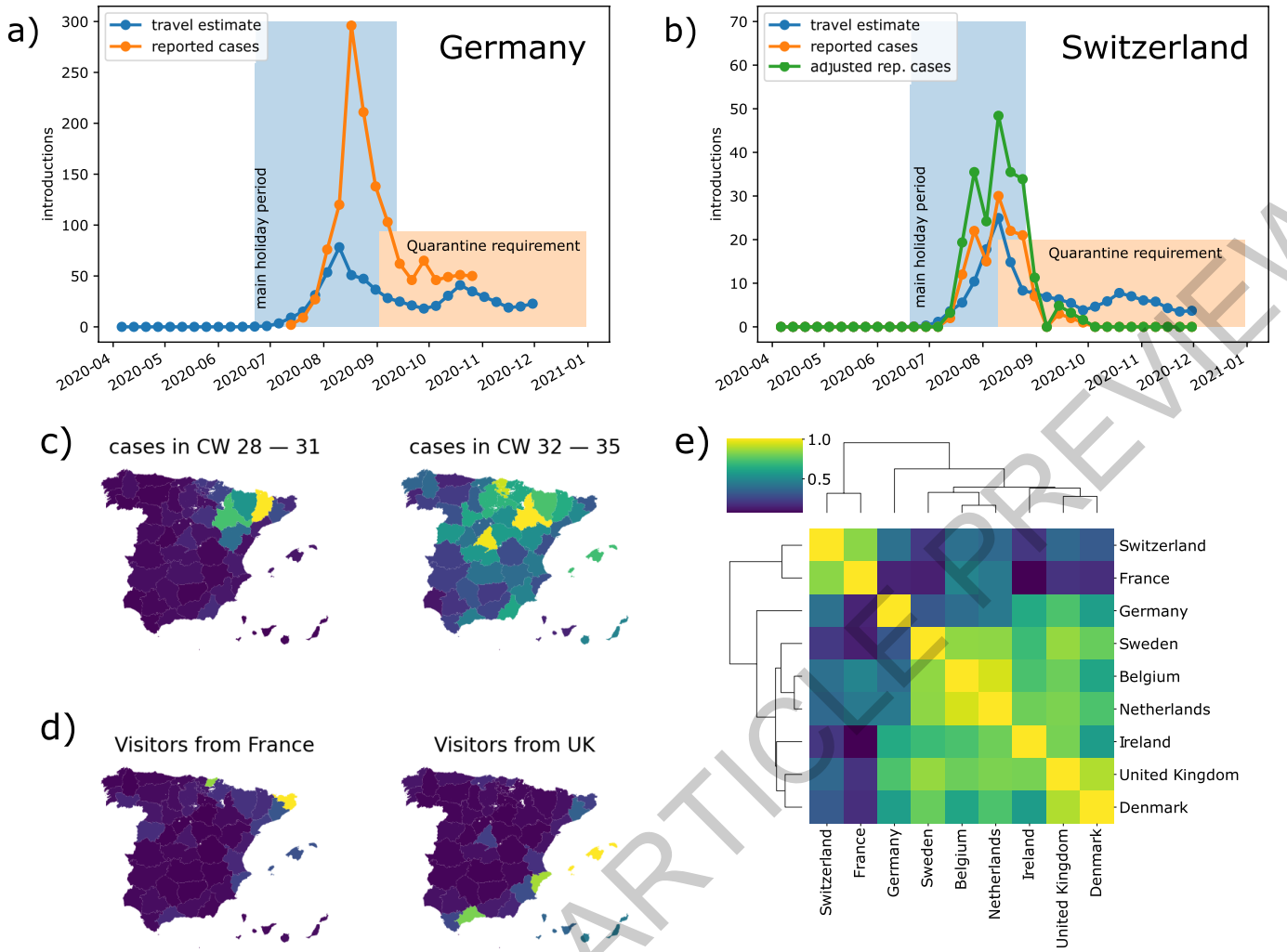


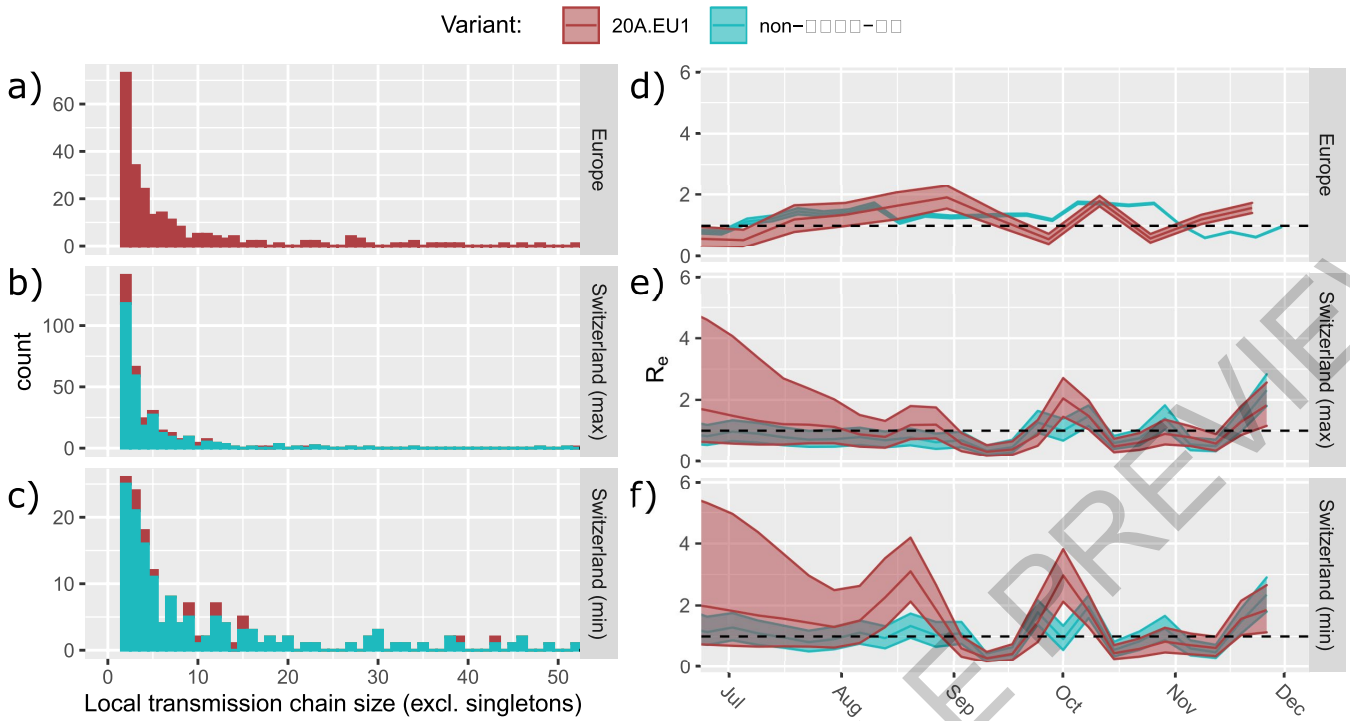

**Extended Data Fig. 5 | Incidence in various countries over the summer.** **A**: Spain and Belgium had relatively higher incidence from the start of July compared with other countries in Europe. **B**: The estimated total number of EU1 cases (red) outside of Spain (countries as in A) surpasses the cases in Spain in September.



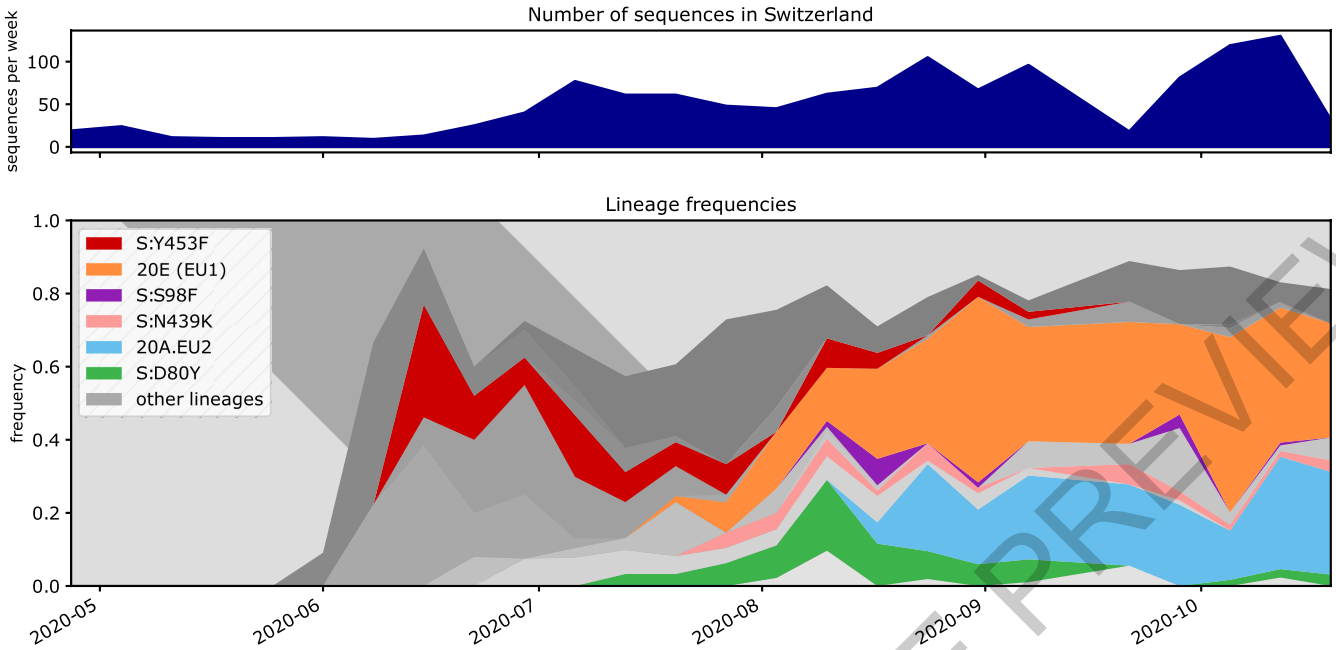

**Extended Data Fig. 6 | Rescaled predictions by the import model match observed frequency trajectories.** In most countries, observations of 20E (EU1) increased in July 2020 and reached a plateau or a slower increase by Oct 2020. Predictions by the import model need to be scaled (see legend) to match the observed frequencies by a factor between 1.2 and 11 (see main text for discussion). Fluctuations on short time scales in the observed frequency of 20E (EU1) are likely due to sampling and dynamics of local outbreaks. Observed frequencies are subject to variable reporting delays.




**Extended Data Fig. 7 | Reported and estimated introductions of 20E (EU1) to Germany (a) and Switzerland (b) and (c-e) incidence in Spain by province and similarity in Spanish province travel destination of selected European countries.** Travel estimate is estimated introductions from Spain based on incidence and roaming data. Reported cases are cases with a suspected origin in Spain as reported by the RKI (Robert-Koch Institute, 2020) and the Federal Office of Public Health (FOPH), for Germany and Switzerland, respectively. In Switzerland the adjusted rep. cases accounts for the fact that 37% of case reports lack exposure information. (**c**) Incidence in Spain in early and mid-summer. (**d**) Distributions of visitors from Spain from different countries. (**e**) Similarities of destinations in Spain among visitors from different countries in calendar weeks 28-35.




**Extended Data Fig. 8 | Phylodynamic analysis of the spread of the 20E (EU1)-variant across Europe (top panels: a & d) and in Switzerland (bottom panels: b, c, e, f).** **(a-c)** The size of putative transmission chains caused by introductions into Europe and Switzerland. Not shown are the number of singletons, which are introductions with no evidence of onward transmission. In Switzerland, these are shown under two extreme definitions of an introduction (min/max; see Methods). Depending on the min/max definition of introductions, there were between 14 or 236 singletons of 20E (EU1) (41 or 81% of all 20E (EU1) introductions) and 62 or 1089 non 20E (EU1) (30 or 79% of all non-20E (EU1) introductions). In Europe, we see 206 20E (EU1) singletons (46% of all 20E (EU1) introductions). There were also a small number of larger transmission chains including more than 53 transmissions (20 across all data sets) which are not shown in the histograms. **(d-e)** The effective reproductive number estimated for 20E (EU1) (red) and the non-20E (EU1) variants (blue). In Switzerland, this is done for the two extreme definitions of an introduction. For Europe, non-20E (EU1) $R_e$ estimates were generated from case numbers. While there is little data to inform estimates of $R_e$ for 20E (EU1) in July and it differs little from the prior, there is some evidence that 20E (EU1) was growing faster than other variants in August. However, systematic differences in ascertainment in travel associated cases might confound this inference. From mid-September, $R_e$ of 20E (EU1) is largely statistically indistinguishable from that of other variants. Shaded areas indicate 95% HPD regions. Notably, the peak in August in the Swiss analysis is larger under the 'min' definition (f) than under the 'max' definition (e), consistent with a more conservative definition of a cluster which would then require more onward transmission.




**Extended Data Fig. 9 | Lineages found in a Swiss-focused Nextstrain build.** A lineage is defined as a node present in the tree after the cut-off date of 1 May 2020 with at least 10 Swiss sequences as children. Clusters discussed in this manuscript are labelled. Lineages are shown as the proportion of the total number of sequences per week in Switzerland. Striped space in the bottom graph represents lineages with most recent common ancestors dating back prior to 1 May 2020 and lineages that do not contain at least 10 Swiss sequences.

**Extended Data Table 1 | Representative mutations of 20E (EU1) (the focus of this study) and other notable variants**

| Variant | Lineage | Representative | |
|---|---|---|---|
| 20E (EU1) | B.1.177 | C22227T, C28932T, G29645T | A222V |
| 20A.EU2 | B.1.160 | C4543T, G5629T, G22992A | S477N |
| S:S98F | B.1.221 | C21855T, A25505G, G25996T | S98F |
| S:D80Y | B.1.367 | C3099T, G21800T, G27632T | D80Y |
| S:N439K | B.1.258 | T7767C, C8047T, C22879A | N439K |

When a lineage definition matches the variant definition, it is given in column 2 (Rambaut et al., 2020[16]).

nature research

Corresponding author(s): Emma B Hodcroft

Last updated by author(s): Nov 21, 2020

# Reporting Summary

Nature Research 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 Research 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<br>*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<br>*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 | *Provide a description of all commercial, open source and custom code used to collect the data in this study, specifying the version used OR state that no software was used.* |
| Data analysis | For phylogenetic analyses we use Nextstrain's augur and auspice packages:<br>https://github.com/nextstrain/augur<br>https://github.com/nextstrain/auspice<br>The code used to run they exact phylogenetic builds is at:<br>https://github.com/emmahodcroft/ncov_cluster<br>For all other analyses and figure plotting we used custom python code that can be found in:<br>https://github.com/emmahodcroft/cluster_scripts |

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 Research 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 list of figures that have associated raw data
- A description of any restrictions on data availability

We provide a table of all accession numbers for the sequence data used from which all raw data can be generated for phylogenetic and sequence analysis. Raw data for the lentiviral experiments can be found at: https://github.com/jbloomlab/A222V-Spike/

# 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 | Not applicable. We used all SARS-CoV-2 samples available on GISAID until 11th November, except for the exclusions outlined below. |
| Data exclusions | We excluded samples that are excluded as part of the official Nextstrain.org builds for divergence and quality control issues as listed in: https://github.com/nextstrain/ncov/blob/master/defaults/exclude.txt<br>We also exclude all samples without a complete date.<br>We outline a few more specific exclusions within the manuscript. |
| Replication | All replications were successful; please see manuscript for details of replicates. |
| Randomization | N/A |
| Blinding | N/A |

# 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 |

## Eukaryotic cell lines

Policy information about cell lines

| | |
|---|---|
| Cell line source(s) | 293T-ACE2 cells (BEI NR-52511) |
| Authentication | The 293T-ACE2 cells are the original source for those available as BEI Resources NR-52515 (https://www.beiresources.org/Catalog/cellBanks/NR-52511.aspx). ACE2 expression was validated by flow cytometry. |
| Mycoplasma contamination | *Confirm that all cell lines tested negative for mycoplasma contamination OR describe the results of the testing for mycoplasma contamination OR declare that the cell lines were not tested for mycoplasma contamination.* |

| | |
|---|---|
| Commonly misidentified lines<br>(See ICLAC register) | *Name any commonly misidentified cell lines used in the study and provide a rationale for their use.* |