High prevalence of SARS-CoV-2 B.1.1.7 (UK variant) and the novel B.1.5.2.5 lineage in Oyo State, Nigeria

Egon A. Ozer[1*], Lacy M. Simons[1*], Olubusuyi M. Adewumi[2,3*], Adeola A. Fowotade[3,4], Ewean C. Omoruyi[4], Johnson A. Adeniji[2,3], Taylor J. Dean[1], Babafemi O. Taiwo[1], Judd F. Hultquist[1], Ramon Lorenzo-Redondo[1,#]

[1] Department of Medicine, Division of Infectious Diseases, Northwestern University Feinberg School of Medicine, Chicago, IL, USA.

[2] Department of Virology, College of Medicine, University of Ibadan, Ibadan, Nigeria.

[3] Infectious Disease Institute, College of Medicine, University of Ibadan, Ibadan, Nigeria.

[4] Biorepository and Clinical Virology Laboratory, College of Medicine, University College Hospital, University of Ibadan, Ibadan, Nigeria.

[*] These authors contributed equally to this work.

[#] Corresponding author. Contact e-mail: ramon.lorenzo@northwestern.edu

ABSTRACT

The spread of SARS-CoV-2, the virus that causes coronavirus disease 2019 (COVID-19), has resulted in a global pandemic that has claimed the lives of millions of people. Genomic surveillance of the virus has proven to be a critical tool for tracking the emergence and spread of variants with increased transmission or immune evasion potential. Despite the global distribution of infection, differences in viral genomic surveillance capabilities between countries and regions have resulted in gaps in our understanding of the viral population dynamics underlying the pandemic. Nigeria, despite having the largest population of any country in Africa, has had relatively little SARS-CoV-2 sequence data made publicly available. In this study, we report the whole-genome sequences of 74 SARS-CoV-2 isolates collected from individuals in Oyo State, Nigeria over the first two weeks of January 2021. Forty-six of the isolates belong to the B.1.1.7 “UK variant” lineage. Comparison to available regional and global sequences suggest that the B.1.1.7 isolates in Nigeria are primarily monophyletic, possibly representing a singular successful introduction into the country. The majority of the remaining isolates (17 of 74) belong to the B.1.525 lineage, which contains multiple spike protein mutations, including the E484K mutation associated with potential immune escape. Indeed, Nigeria has the highest reported frequency of this lineage despite its relative rarity worldwide. Phylogenetic analysis of the B.1.525 isolates in this study relative to other local and global isolates suggested a recent origin and rapid expansion of this lineage in Nigeria, with the country serving as a potential source for this lineage in other outbreaks. These results demonstrate the importance of genomic surveillance for identifying SARS-CoV-2 variants of concern in Nigeria and in other undersampled regions across the globe.

KEYWORDS

SARS-CoV-2, COVID-19, Phylogenetics, Lineage, Variant of concern, Viral genotype, Whole genome sequencing, Global health

INTRODUCTION

A little over a year after its emergence in the Hubei province of China, the continued spread of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) across the globe has sparked a worldwide health crisis (1-3). Despite the relatively low mutation rate of this virus, its high prevalence in the human population globally has allowed it to diversify quickly (4, 5). Identification and tracking of these mutations through whole genome sequencing efforts have been critical to identifying routes of transmission, mapping outbreaks across communities over time, and characterizing new variants that may change the virological or clinical aspects of the disease (6-9). For example, during the spring and summer of 2020, a novel mutation in the viral Spike protein, D614G, was identified in association with higher viral loads in patient upper airways (10, 11). The rapid expansion of this variant in diverse communities across the world coupled with virological assays *in vitro* and *in vivo*, suggest that this variant is more transmissible (12-15). Indeed, this mutation is nearly fixed in the SARS-CoV-2 population today (11).

Continued surveillance efforts have since identified a number of 'variants of concern' that have been associated with rapid expansion in their local communities and spread to other countries. Using the Pangolin nomenclature, these include the B.1.1.7 lineage originally identified in the United Kingdom, the B.1.351 variant originally identified in South Africa, and the P.1 and P.2 variants originally identified in Brazil (16-21). The first concern is that these viruses may prove to be more transmissible and could render current public health standards less effective at containing spread of the virus (22-24). The second is that changes to the viral Spike protein may render current formulations of the vaccine less efficacious and/or confer increased capacity for re-infection (25-27). Finally, viral mutations may impact the effectiveness of therapeutic monoclonocal antibodies(28). Addressing these concerns requires continued surveillance of SARS-CoV-2 genetic diversity worldwide.

While many countries have developed extensive genetic surveillance and reporting systems, several regions across the globe remain critically undersampled (29). For example, several countries in Africa have reported only a handful of sequences relative to their cumulative case counts (30, 31). This need is even more acute given the recent emergence of several variants of global concern. Significantly, Nigeria, which has the highest total population in the continent, has only a small number of reported SARS-CoV-2 sequences, limiting our understanding of the viral population structure in the country and continent. Through mid-February of 2021, the majority of the approximately 300 SARS-CoV-2 genome sequences publicly available from Nigeria have been generated through the African Centre of Excellence for Genomics of Infectious Disease (ACEGID), which also serves other African countries (www.acegid.org). However, given Nigeria's status as an epicenter of commerce and travel in Africa, undetected expansion of a more infectious, virulent or immune-escape variant in Nigeria will have major repercussions for the entire continent. More consistent and higher volume of sample collection and sequencing is required in Nigeria to strengthen the public health value of serosurveillance.

To better understand the current SARS-CoV-2 population structure in Nigeria and add samples to this region, we set out to perform whole genome sequencing of SARS-CoV-2 viruses originating in Oyo state, Nigeria. SARS-CoV-2 whole genome sequencing was performed on 74 nasopharyngeal specimens collected from COVID-19 patients in Oyo state in January 2021. Phylogenetic analysis revealed two primary lineages of virus circulating in the region: B.1.1.7 (*i.e.*, the 'U.K. variant') and B.1.525, an emerging variant of interest. Both lineages are relatively monophyletic, suggesting singular introductions sometime in the fall of 2020. These data help to fill in a significant gap in SARS-CoV-2 genomic surveillance in Nigeria and identify two lineages of note currently dominating the epidemic in the region.

RESULTS

Specimen characteristics

We obtained 80 specimens from COVID-19 patients in Oyo state, Nigeria collected between January 2 and January 14, 2021. These specimens consisted of residual nasopharyngeal and oropharyngeal swabs that tested positive for SARS-CoV-2 by quantitative reverse transcriptase PCR (qRT-PCR) diagnostic testing in the Biorepository and Clinical Virology Laboratory at the University College Hospital College of Medicine at University of Ibadan. The qRT-PCR cycle threshold (Ct) values of these specimens obtained in the clinical laboratory for N gene using either DaAn Gene or BGI detection kits ranged from 13.46 to 33.11 and 18.51 to 37.10 for ORF1ab when using DaAn gene kit. The average age of the patients was 42 years (range [11-84]) and the gender distribution was 50% female and 50% male. RNA was extracted from each sample and the level of viral RNA was recertified by one-step qRT-PCR (CDC assay, RNaseP and N1 primer set)(32). Samples with high quality RNA (RNaseP Ct value >35) and sufficient viral RNA for whole genome sequencing (N1 Ct value <32) were reverse transcribed into complementary DNA (cDNA). The SARS-CoV-2 genome was subsequently amplified by multiplex PCR using the ARTIC protocol (primer set version 3) and deep sequenced on the Illumina platform (33, 34). The minimum threshold for base calling was 10 reads with 90% coverage required to report the whole genome sequence. Of the 80 specimens, 6 failed to yield a final consensus sequence due to insufficient genetic material, insufficient purity after barcoding, or inadequate read coverage after sequencing; these samples were excluded from further analysis. The final 74 complete SARS-CoV-2 genomes were deposited in the publicly available GISAID database (Supplemental Table 1) and subjected to phylogenetic analysis (35, 36).

The two dominant lineages in Oyo state are the lineage of concern B.1.1.7 and the novel lineage of interest B.1.525.

We first performed phylogenetic reconstruction using maximum likelihood (ML) phylogenetic analysis of the 74 SARS-CoV-2 genomes in this study using IQ-Tree v2.0.5. The dates of sample collection were subsequently integrated to build a temporal tree using TreeTime v0.7.6 (**Figure 1**). Most of the specimens from Oyo state (88%) belonged to one of two main clusters with very strong support at their base nodes (support: >97% aLRT and >97% UFboot). We used the Pangolin method of classification to identify the lineages of the clusters

observed. The most abundant lineage (46 out of 74 sequences) was B.1.1.7. This lineage of concern is associated with the Spike N501Y mutation and was first identified in the U.K. in the fall of 2020 (20). The second most common lineage (17 of 74 sequences) was a lineage of interest, designated B.1.525. This lineage contains the Spike E484K, Q677H, and F888L mutations as well as three in-frame deletions shared with B.1.1.7, including Spike 69-70del (cov-lineages.org). The remaining sequences all belonged to other common B.1.1 lineages, except for 2 sequences from lineage A. Viruses from lineage A viruses spread throughout the world early in the pandemic, but are now relatively rare globally, accounting for only 0.3% of isolates collected in January 2021 (nextstrain.org).

Comparison with other publicly available Nigerian isolate sequences

To better determine if these sequences in Oyo state were representative of Nigeria as a whole, we downloaded all available sequences sourced from Nigeria in the GISAID database as of February 14th (266 sequences). We followed the same approach as above and performed ML phylogenetic reconstruction to estimate the evolutionary relationships with the other available sequences. We also performed ancestral reconstruction of the most likely sequences at internal nodes as well as at transition points between geographical locations to better examine relationships between the isolates found in different states in Nigeria. This analysis confirmed a distribution of the viral populations throughout the country similar to what was observed in Oyo state. Of all recently reported sequences (January 1, 2021 – February 14, 2021), we observe that 68% of the sequences belong to the B.1.1.7 lineage while 17% belong to the B.1.525 lineage (**Figure 2**). This includes sequences from Lagos state, Osun state, Abuja, and other sequences from Oyo state collected outside this study. Overall, this result suggests that the distribution of lineages observed in this study is broadly relective of the recent distribution of lineages in Nigeria. It furthermore confirms the recent increase in prevalence of these two lineages in Nigeria, the emergence of which coincided with an increase in overall case count (37).

Nigerian viral populations compared to global sequences and B.1.1.7 analysis

To better understand these data in the context of the global epidemic, we selected 4000 randomly sampled global sequences from GISAID and repeated these analyses. We confirmed the clustering of the lineages as observed above (**Figure 3**). Furthermore, these data reveal a monophyletic clade of most of the B.1.1.7 lineage sequences in Nigeria, suggesting either a single successful introduction and outbreak of this specific B.1.1.7 sub-cluster or multiple, closely linked introductions compatible with the observed founder effect, i.e. introduction of a subset of mutations and decrease in lineage diversity as observed in B.1.1.7 sequences in Nigeria. To confirm the B.1.1.7 sub-clustering in Nigeria, we performed a ML phylogenetic reconstruction of all B.1.1.7 sequences obtained in this study alongside 4000 randomly sampled B.1.1.7 sequences from across the globe (**Figure 4**). Additionally, we performed Bayesian inference of all B.1.1.7 sequences available from Nigeria with a smaller set of 500 randomly sampled Global B.1.1.7 sequences (**Figure 5**). These analyses both confirmed strong support for a monophyletic sub-cluster that included almost the entirety of the B.1.1.7 sequences from this study. Given these results, we estimated a time to the most recent common ancestor (TMRCA) for this B.1.1.7 sub-cluster of October 28, 2020 [95% Highest Posterior Density (HPD) interval: October 16 - November

3, 2020]. Together, these results suggest that the expansion and dominance of B.1.1.7 in Nigeria can be traced back to a single introduction or multiple, closely timed introductions in the fall of 2020.

Analysis of the recently identified B.1.525 lineage in Nigeria

We also completed a detailed analysis of the second most prevalent lineage observed in our dataset, B.1.525. This lineage is not highly prevalent on the global scale but was recently defined as a lineage of international significance (38). At the time of this analysis, only 159 sequences were available in GISAID from this lineage, 17 coming from this dataset. As of February 2021, Nigeria had highest frequency of the B.1.525 lineage in any country worldwide, accounting for roughly 25% of sequences (38). Referring to our previous ML analysis of the sequences from Oyo state alongside 4000 randomly selected global sequences (**Figure 3**), we observe strong support for a monophyletic clade of the B.1.525 lineage in Nigeria, suggestive of a single introduction followed by rapid expansion (**Figure 3**). Using a Bayesian approach, we next analyzed all 159 B.1.525 sequences available in GISAID alongside a smaller subset of B.1 sequences to help root the tree (the only 2 B.1 sequences present in our dataset and 10 randomly selected from GISAID). The estimated TMRCA was October 27, 2020 [95% HPD interval: September 27 - November 21, 2020], similar to the TMRCA estimated for the B.1.1.7 lineage. This analysis suggests that this lineage may have expanded principally in Nigeria, making this country the most likely source of a majority of sequences observed elsewhere in the world (**Figure 6**). That being said, we cannot conclude that the B.1.525 lineage originated in Nigeria due to low sampling and the observation that two sequences from Spain root outside of the node with the most statistical support corresponding to TMRCA for the B.1.525 lineage in Nigeria. In sum, our analyses are all consistent with a single introduction or multiple, closely linked introductions of both the B.1.1.7 and B.1.525 lineages into Nigeria in the fall of 2020. These introductions were followed by expansion of each population to become the predominant two lineages in the country by early 2021.

No significant differences are observed in the Ct values of the diagnostic PCR tests by clade

Finally, we tested for differences in the viral load observed at diagnosis, as represented by cycle threshold (Ct) value, by viral lineage. Using a linear model to control for patient age and gender, we compared the Ct value measured for the N1 probe for all 74 diagnostic specimens batched by lineage (**Figure 7**). We did not observe any statistically significant differences between the two main lineages (B.1.1.7 and B.1.525), and the remaining lineages lacked sufficient representation for meaningful comparison. Although a larger dataset that controls for more confounders will be needed to robustly test for differences, this result suggests there are comparable average viral loads in the upper respiratory tract between these two lineages at the time of patient diagnosis.

DISCUSSION

Together, these results indicate the dominance of two distinct viral populations in Oyo state and in Nigeria more broadly. The B.1.1.7 lineage, which was first reported on December 21, 2020 in the U.K. and first sampled on September 20, 2020, appears to have become dominant within Oyo state and throughout the other sampled

states in Nigeria by early January 2021. While this variant of concern has not been found to influence efficacy of neutralizing antibodies produced upon vaccination, it has been suggested to be more transmissible and can interfere with diagnostic tests probing over the Spike deletion (20, 39-41). These data should be considered in updating the diagnostic and prevention strategies in Nigeria. More broadly, these data increase the call for international guidance on establishing public health policy for countries with high prevalence of variants.

Another significant observation of this study is the previously unreported increase of the B.1.525 lineage in Nigeria. The prevalence of the B.1.525 lineage in Nigeria was the highest observed in any country, with phylodynamic data suggesting Nigeria may have served as the origin for several outbreaks of this lineage. Critically, this lineage contains the Spike E484K mutation, which some studies have suggested may impact immune recognition and vaccine efficacy (42-47). It is imperative to monitor this lineage and better understand its impact on vaccine response as vaccination expands in Nigeria and across the region.

While these analyses suggest both lineages arose from either singular or multiple, closely linked introductions to Nigeria, overall lack of sampling across this region remains a significant confounder and more data will be needed to further test these hypotheses. Comparing the estimated TMRCA for both lineages with the daily incidence reported by the NCDC (**Figure 8**), we note that the possible introduction or appearance of these lineages coincides with some of the lowest daily incidence numbers since the beginning of the pandemic. This observation suggests the possibility that both lineages benefited from founder-effects that boosted overall prevalence of each as case numbers rose. Less than a month prior to the estimated TMRCA for these lineages, Nigeria reopened international airports to regular air traffic (September 5, 2020). This step was accompanied by strong measures to prevent SARS-CoV-2 importation, including requiring a negative PCR test within 96 hours of boarding an in-bound flight, screening for COVID-19 symptoms prior to boarding the flight, self-quarantine for 7 days after arrival in Nigeria, and a mandatory second PCR test on day 7 of arrival. Despite these rigorous public health measures, these data are suggestive of a model wherein these lineages were seeded from international travel followed by local expansion.

The successful introduction and/or spread of these lineages in a period of very low incidence, together with their possible fitness advantages, could have facilitated the observed dominance of these lineages in Nigeria. Due to the suggested higher infectivity rate of B.1.1.7 and suggested decreased antibody elicited neutralization capacity against E484K variants, it is possible that these lineages could be in part responsible for the significant increase of daily incidence observed in Nigeria in late 2020. That being said, cases increased significantly across the globe regardless of variant prevalence, suggesting additional factors are at play, such as increased socialization at the end of the year. Furthermore, the introduction of these linegaes at a time of low prevelenace will likely have contributed to their expansion through founder effect. Regardless, additional sequencing and surveillance across the region, and elucidation of their biological and clinical significance, will be required to better understand and cope with emerging variants on the global scale.

METHODS

Viral RNA extraction

Viral RNA was extracted from clinical specimens utilizing the QIAamp Viral RNA Minikit (Qiagen, cat. no. 52906). Clinical testing for SARS-CoV-2 presence was performed by quantitative reverse transcription and PCR (qRT-PCR) with the CDC 2019-nCoV RT-PCR Diagnostic Panel utilizing the N1 probe in SARS-CoV-2 and RP probes for sample quality control as previously described (IDT, cat. no. 10006713). All specimens that failed to amplify the RP housekeeping gene were excluded from this study. All specimens with an N1 probe cycle threshold (Ct) less than or equal to 35 were considered positive and included in this study. RT-PCR was repeated in a random selection of specimens to validate Ct values obtained by the clinical diagnostic laboratory. Ct values from the N1 probes were used in all subsequent analyses.

cDNA synthesis and viral genome amplification

cDNA synthesis was performed with SuperScript IV First Strand Synthesis Kit (ThermoFisher, cat. no. 18091050) using 11 µl of extracted viral nucleic acids and random hexamers according to manufacturer's specifications. Direct amplification of the viral genome cDNA was performed in multiplexed PCR reactions to generate ~400 bp amplicons tiled across the genome. The multiplex primer set, comprised of two non-overlapping primer pools, was created using Primal Scheme and provided by the Artic Network (version 3 release). PCR amplification was carried out using Q5 Hot Start HF Taq Polymerase (NEB, cat. no. M0493L) with 5 µl of cDNA in a 25 µl reaction volume. A two-step PCR program was used with an initial step of 98 °C for 30 s, then 35 cycles of 98 °C for 15 s followed by five minutes at 64 °C. Separate reactions were carried out for each primer pool and validated by agarose gel electrophoresis alongside negative controls. Each reaction set included positive and negative amplification controls and was performed in a space physically separated for pre- and post-PCR processing steps to reduce contamination. Amplicon sets for each genome were pooled prior to sequencing library preparation.

Sequencing library preparation, Illumina sequencing, and Genome Assembly

Sequencing library preparation of genome amplicon pools was performed using the SeqWell plexWell 384 kit per manufacturer’s instructions. Pooled libraries of up to 96 genomes were sequenced on the Illumina MiSeq using the V2 500 cycle kit. Sequencing reads were trimmed to remove adapters and low quality sequences using Trimmomatic v0.36. Trimmed reads were aligned to the reference genome sequence of SARS-CoV-2 (accession MN908947.3) using bwa v0.7.15. Pileups were generated from the alignment using samtools v1.9 and consensus sequence determined using iVar v1.2.2 (ref PMID: 30621750) with a minimum depth of 10, a minimum base quality score of 20, and a consensus frequency threshold of 0 (i.e. majority base as the consensus).

Phylogenetic analysis

Genome sequences were aligned using MAFFT v7.453 software and manually edited using MEGA v6.06. All Maximum Likelihood (ML) phylogenies were inferred with IQ-Tree v2.0.5 using its ModelFinder function before each analysis to estimate the nucleotide substitution model best-fitted for each dataset by means of Bayesian information criterion (BIC). We assessed the tree topology for each phylogeny both with the Shimodaira–Hasegawa approximate likelihood-ratio test (SH-aLRT) and with ultrafast bootstrap (UFboot) with 1000 replicates each. TreeTime v0.7.6 was used for the assessment of root-to-tip correlation, the estimation of time scaled phylogenies and ancestral reconstruction of most likely sequences of internal nodes of the tree and transitions between geographical locations along branches. TreeTime was run using an autocorrelated molecular clock under a skyline coalescent tree prior. We used the sampling dates of the sequences to estimate the evolutionary rates and determine the best rooting of the tree using root-to-tip regression with least-squares method.

Bayesian time-scaled phylogenetic analyses were performed for the phylogenies of B.1.1.7 and B.1.525 lineages separately using a smaller set (for computational time reasons) of of 500 randomly sampled global sequences from GISAID for B.1.1.7 and all the available sequences for B.1.525. Due to the small number of B.1.525 sequences available, we included B.1 sequences from our Nigerian dataset as well as 10 random B.1 sequences from GISAID to help root the phylogenies. We used BEAST v2.5.2 to estimate the date and location of the most recent common ancestors (MRCA) as well as to estimate the rate of evolution of the virus. BEAST priors were introduced with BEAUTI v2.5.2 including an uncorrelated relaxed molecular clock model with a lognormal distribution of the evolutionary rate, previous estimated evolutionary rates (8x10-4) as the prior for the mean, and a standard deviation of 0.1 after optimization with preliminary runs. We assumed a GTR substitution model with invariant sites, as the best-fitted model obtained with ModelFinder, and a Coalescence Bayesian Skyline to model the population size changes through time. The posterior evolutionary rate was estimated to be 5.9x10-4 for the B.1.1.7 inference and 7.9x10-4 for the B.1.525 analysis (that included B.1 lineages). All the analyses indicated support for a relaxed clock as the average rate coefficient of variation was 0.2 for B.1.1.7 and 0.6 for B.1.525, which indicates a degree of rate autocorrelation among adjacent branches in the tree. Markov chain Monte Carlo (MCMC) runs of at least 100 million states with sampling every 5,000 steps were computed. The convergence of MCMC chains was monitored using Tracer v.1.7.1, ensuring that the effective sample size (ESS) values were greater than 200 for each parameter estimated.

ACKNOWLEDGEMENTS

This research was supported in part through the computational resources and staff contributions provided for the Quest high performance computing facility at Northwestern University, which is jointly supported by the Office of the Provost, the Office for Research, and Northwestern University Information Technology. Funding for this work was provided by: a Northwestern Institute for Global Health Catalyzer Research Fund award (R.L.R. and O.M.A.); a Dixon Translational Research Grant made possible by the generous support of the Dixon Family

Foundation (E.A.O. and J.F.H.); a COVID-19 Supplemental Research award from the Northwestern Center for Advanced Technologies (NUCATS - J.F.H.); a CTSA supplement to NCATS UL1 TR002389 (J.F.H., E.A.O., R.L.R.); a supplement to the Northwestern University Cancer Center P30 CA060553 (J.F.H.); the Gilead Sciences Research Scholars Program in HIV (J.F.H.); the NIH-supported Third Coast CFAR P30 AI117943 (R.L.R., J.F.H.); NIH grant K22 AI136691 (J.F.H.); NIH grant U19 AI135964 (E.A.O.); NIH grant D43 TW009608 (B.O.T., O.M.A.); NIH Fogarty International Center award D43TW009608 (B.O.T.); and through a generous contribution from the Walder Foundation (J.F.H., E.A.O., R.L.R.). The authors also acknowledge the Nigeria Center for Disease Control (NCDC) for providing leadership and infrastructure that facilitated collection of samples included in this study. The funding sources had no role in the study design, data collection, analysis, interpretation, or writing of the report.

AUTHOR CONTRIBUTIONS

Conceptualization: R.L.R., L.M.S., J.F.H., E.A.O., B.O.T. Investigation: O.M.A., A.A.F., E.C.O., J.A.A., T.J.D., R.L.R., L.M.S., J.F.H., E.A.O. Resources: O.M.A., J.F.H., E.A.O. Formal analysis: R.L.R., E.A.O. Supervision: J.F.H., E.A.O. Funding acquisition: J.F.H., E.A.O., R.L.R., B.O.T. Writing: R.L.R., L.M.S., J.F.H., E.A.O., B.O.T., O.M.A.

CONFLICTS OF INTEREST

None.

FIGURE LEGENDS

Figure 1. Phylogenetic analysis of SARS-CoV-2 isolates in Oyo state. a) ML phylogenetic tree of 74 SARS-CoV-2 specimen genomes in Oyo state collected between January 2 and January 14, 2021. All non-zero statistical support values for each branch are indicated. Lineages of interest are indicated and colored. Midpoint rooting was used for representation purposes. **b)** Distribution of the different pangolin lineages found in the Oyo dataset reported here.

Figure 2. Phylogenetic analysis of Oyo state isolates compared to all available Nigerian sequences. ML phylogenetic temporal reconstruction of full genome sequences from Nigeria, including the sequences from this study and all sequences available from Nigeria in GISAID as of February 14th, 2021. Clades corresponding to B.1.1.7 and B.1.525 lineages are indicated. Branches and tips are colored by state; labels corresponding to sequences obtained in this study are colored in red.

Figure 3. Phylogenetic analysis of Nigerian sequences compared to the global pandemic. ML phylogenetic temporal reconstruction of full genome sequences from Nigeria and 4000 randomly sampled global sequences

from GISAID as of February 14th, 2021. Clades corresponding to B.1.1.7 and B.1.525 lineages are indicated. Branches and tips are colored by country.

Figure 4. Phylogenetic analysis of Oyo state B.1.1.7 sequences compared to global B.1.1.7 sequences. ML phylogenetic temporal reconstruction of full genome B.1.1.7 sequences obtained in this study and 4000 randomly sampled B..1.1.7 global sequences from GISAID as of February 14th, 2021. Branches and tips are colored by country.

Figure 5. Phylodynamic tree of Nigeria B.1.1.7 sequences compared to global B.1.1.7 sequences. Maximum clade credibility tree where branch colors represent the most probable geographical location of their descendent node inferred through Bayesian reconstruction of the ancestral state. All full genome B.1.1.7 sequences from Nigeria and 500 randomly sampled B.1.1.7 global sequences from GISAID as of February 14th, 2021 were included in the analysis. The width of the node circles represents their posterior probability.

Figure 6. Phylodynamic tree of the entire B.1.525 lineage. Maximum clade credibility tree where branch colors represent the most probable geographical location of their descendent node inferred through Bayesian reconstruction of the ancestral state. All full genome B.1.525 sequences from this study and from GISAID as of February 14th, 2021 were included in the analysis. We included two B.1 lineage sequences from Nigeria and 10 randomly sampled from GISAID to help root the phylogenies. The branch that leads to the node that represents the B.1.525 MRCA is indicated. The width of the node circles represents their posterior probability.

Figure 7. PCR Cycle threshold (Ct) values of patient diagnostic samples grouped by pangolin lineage assignment. Ct values for the N1 probe set reported at the time of diagnosis were compared between lineage. A linear model was fitted to test for differences in Ct value between lineages. All possible contrasts within the model were performed and corrected for multiple comparisons using Tukey's test. No corrected p-values were found to be statistically significant using 0.05 cut-off.

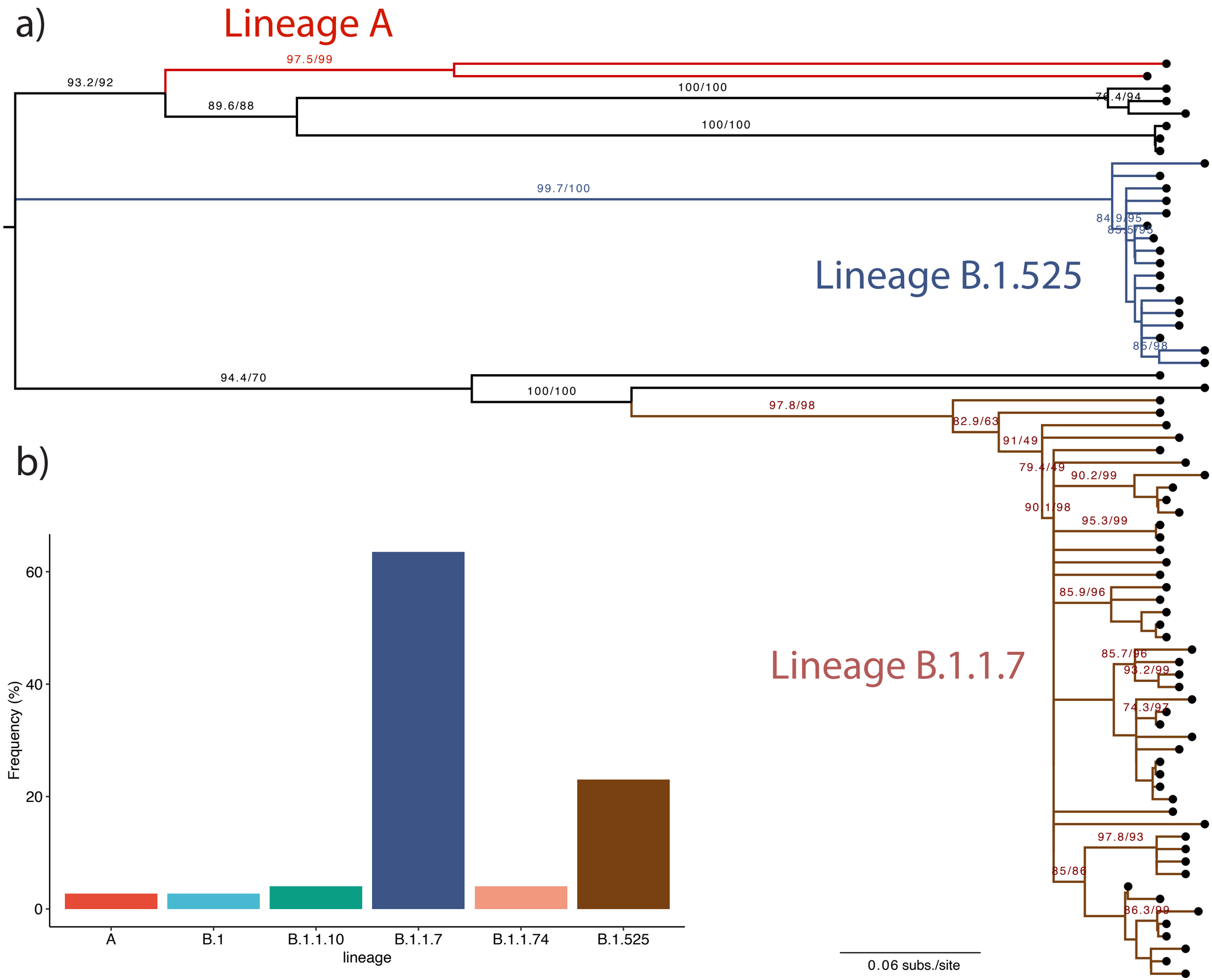
Figure 8. Daily SARS-CoV2 Incidence in Nigeria. Confirmed new cases in Nigeria obtained from Johns Hopkins University Coronavirus resource center (https://coronavirus.jhu.edu/). The TMRCA (solid line) and 95% High Probability Density (HPD) (dashed lines) in Nigeria of B.1.1.7 and B.1.525 lineages estimated using Bayesian methods is indicated.

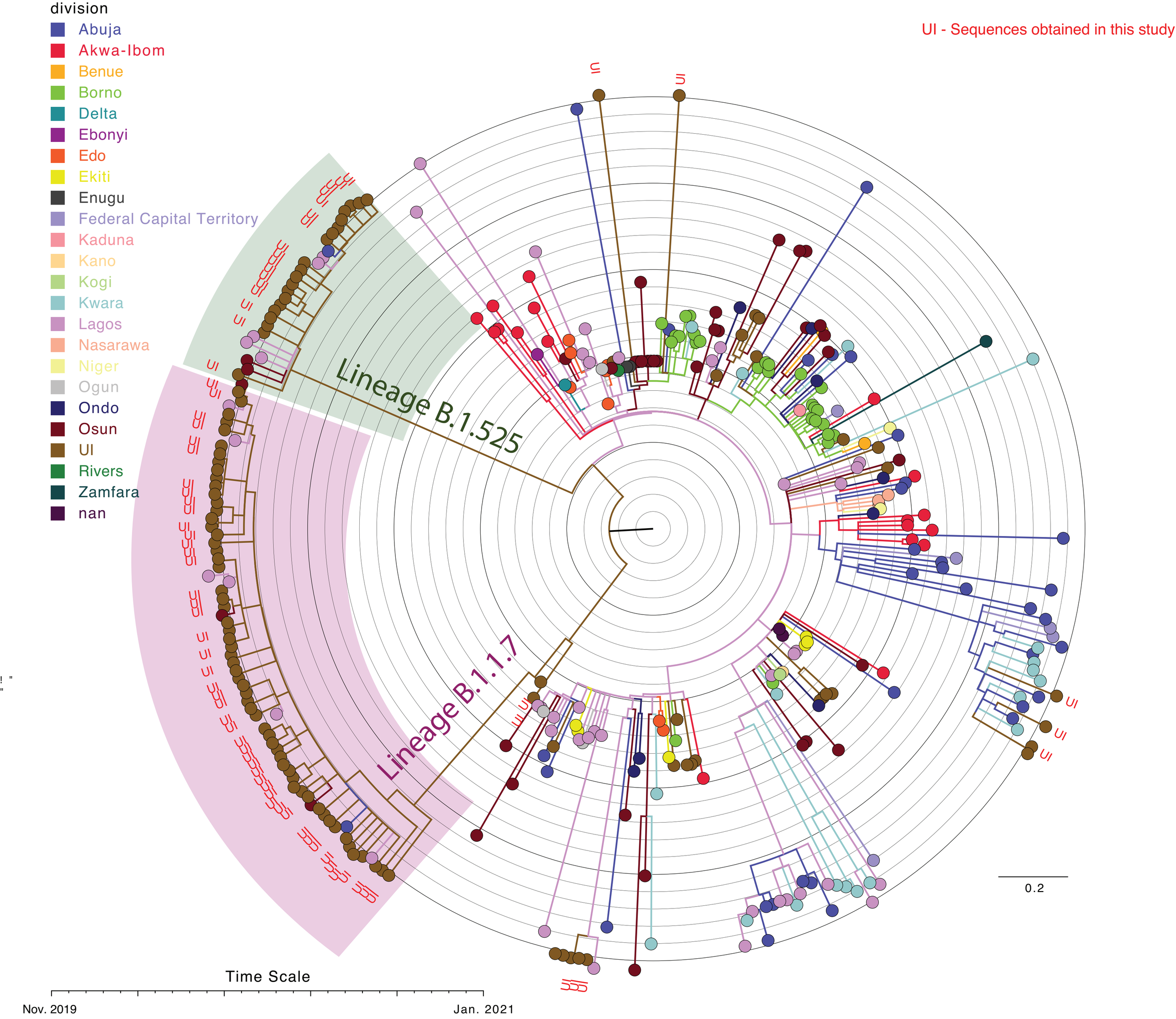
REFERENCES

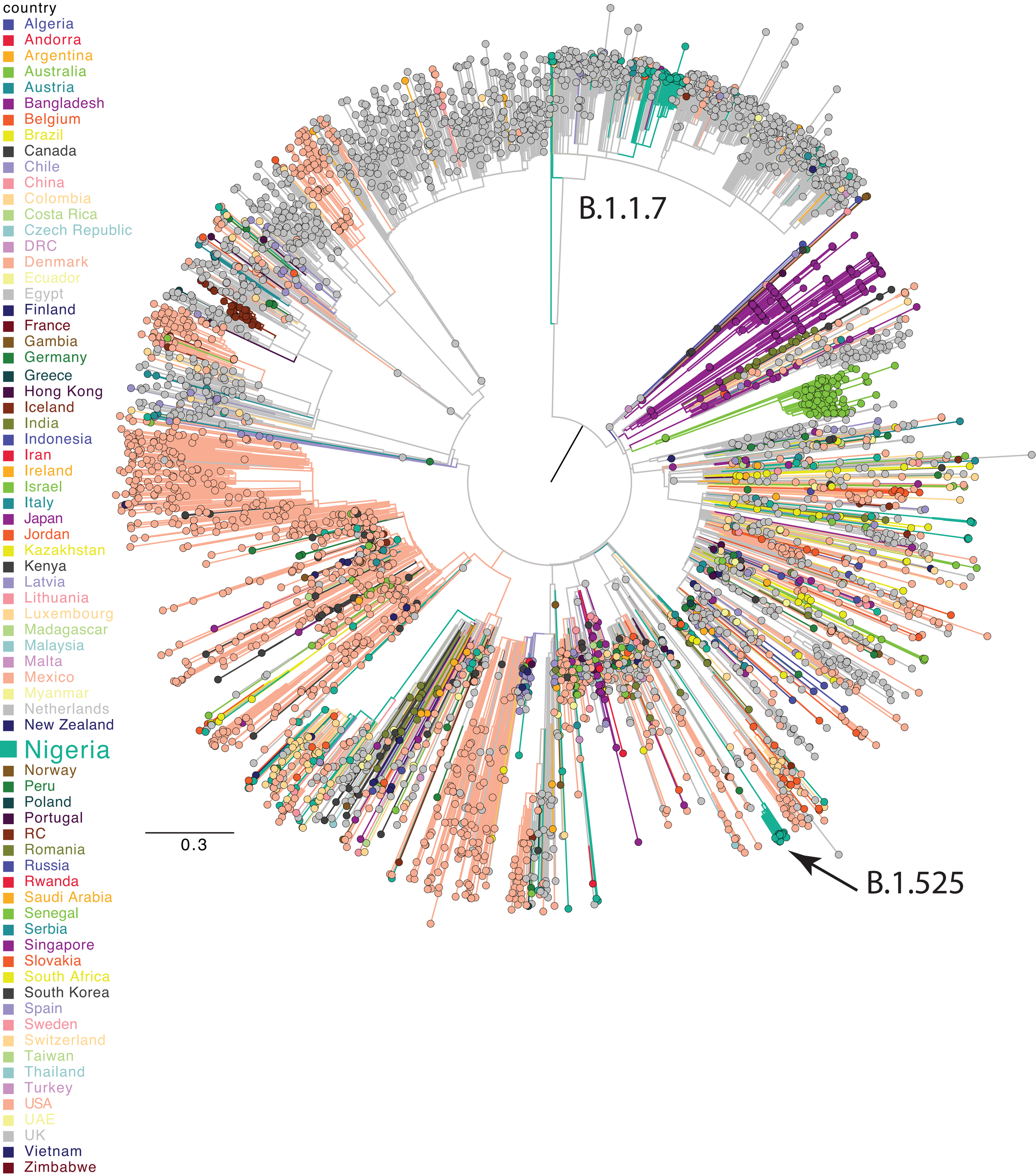
1. Wang C, Horby PW, Hayden FG, Gao GF. A novel coronavirus outbreak of global health concern. Lancet. 2020;395(10223):470-3.
2. Zhu N, Zhang D, Wang W, Li X, Yang B, Song J, et al. A Novel Coronavirus from Patients with Pneumonia in China, 2019. N Engl J Med. 2020;382(8):727-33.
3. Zhou P, Yang XL, Wang XG, Hu B, Zhang L, Zhang W, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature. 2020;579(7798):270-3.
4. van Dorp L, Acman M, Richard D, Shaw LP, Ford CE, Ormond L, et al. Emergence of genomic diversity and recurrent mutations in SARS-CoV-2. Infect Genet Evol. 2020;83:104351.
5. Li X, Wang W, Zhao X, Zai J, Zhao Q, Li Y, et al. Transmission dynamics and evolutionary history of 2019-nCoV. J Med Virol. 2020;92(5):501-11.
6. Meredith LW, Hamilton WL, Warne B, Houldcroft CJ, Hosmillo M, Jahun AS, et al. Rapid implementation of SARS-CoV-2 sequencing to investigate cases of health-care associated COVID-19: a prospective genomic surveillance study. Lancet Infect Dis. 2020;20(11):1263-71.
7. Dearlove B, Lewitus E, Bai H, Li Y, Reeves DB, Joyce MG, et al. A SARS-CoV-2 vaccine candidate would likely match all currently circulating variants. Proc Natl Acad Sci U S A. 2020;117(38):23652-62.
8. Lu J, du Plessis L, Liu Z, Hill V, Kang M, Lin H, et al. Genomic Epidemiology of SARS-CoV-2 in Guangdong Province, China. Cell. 2020;181(5):997-1003 e9.
9. Gonzalez-Reiche AS, Hernandez MM, Sullivan MJ, Ciferri B, Alshammary H, Obla A, et al. Introductions and early spread of SARS-CoV-2 in the New York City area. Science. 2020;369(6501):297-301.
10. Lorenzo-Redondo R, Nam HH, Roberts SC, Simons LM, Jennings LJ, Qi C, et al. A clade of SARS-CoV-2 viruses associated with lower viral loads in patient upper airways. EBioMedicine. 2020;62:103112.
11. Korber B, Fischer WM, Gnanakaran S, Yoon H, Theiler J, Abfalterer W, et al. Tracking Changes in SARS-CoV-2 Spike: Evidence that D614G Increases Infectivity of the COVID-19 Virus. Cell. 2020;182(4):812-27 e19.
12. Hou YJ, Chiba S, Halfmann P, Ehre C, Kuroda M, Dinnon KH, 3rd, et al. SARS-CoV-2 D614G variant exhibits efficient replication ex vivo and transmission in vivo. Science. 2020;370(6523):1464-8.
13. Li Q, Wu J, Nie J, Zhang L, Hao H, Liu S, et al. The Impact of Mutations in SARS-CoV-2 Spike on Viral Infectivity and Antigenicity. Cell. 2020;182(5):1284-94 e9.
14. Volz E, Hill V, McCrone JT, Price A, Jorgensen D, O'Toole A, et al. Evaluating the Effects of SARS-CoV-2 Spike Mutation D614G on Transmissibility and Pathogenicity. Cell. 2021;184(1):64-75 e11.
15. Plante JA, Liu Y, Liu J, Xia H, Johnson BA, Lokugamage KG, et al. Spike mutation D614G alters SARS-CoV-2 fitness. Nature. 2020.
16. Tegally H, Wilkinson E, Lessells RJ, Giandhari J, Pillay S, Msomi N, et al. Sixteen novel lineages of SARS-CoV-2 in South Africa. Nat Med. 2021.
17. Leung K, Shum MH, Leung GM, Lam TT, Wu JT. Early transmissibility assessment of the N501Y mutant strains of SARS-CoV-2 in the United Kingdom, October to November 2020. Euro Surveill. 2021;26(1).
18. Faria NR, Claro IM, Candido D, Franco LAM, Andrade PS, Coletti TM, et al. Genomic characterisation of an emergent SARS-CoV-2 lineage in Manaus: preliminary findings. Jan 12, 2021. Available from: https://virological.org/t/genomic-characterisation-of-an-emergent-sars-cov-2-lineage-in-manaus-preliminary-findings/586.
19. Volz E, Mishra S, Chand M, Barrett JC, Johnson R, Geidelberg L, et al. Transmission of SARS-CoV-2 Lineage B.1.1.7 in England: Insights from linking epidemiological and genetic data. medRxiv. 2021:2020.12.30.20249034.
20. Rambaut A, Loman N, Pybus O, Barclay W, Barrett J, Carabelli A, et al. Preliminary genomic characterisation of an emergent SARS-CoV-2 lineage in the UK defined by a novel set of spike mutations. Dec 18, 2020. Available from: https://virological.org/t/preliminary-genomic-characterisation-of-an-emergent-sars-cov-2-lineage-in-the-uk-defined-by-a-novel-set-of-spike-mutations/563.

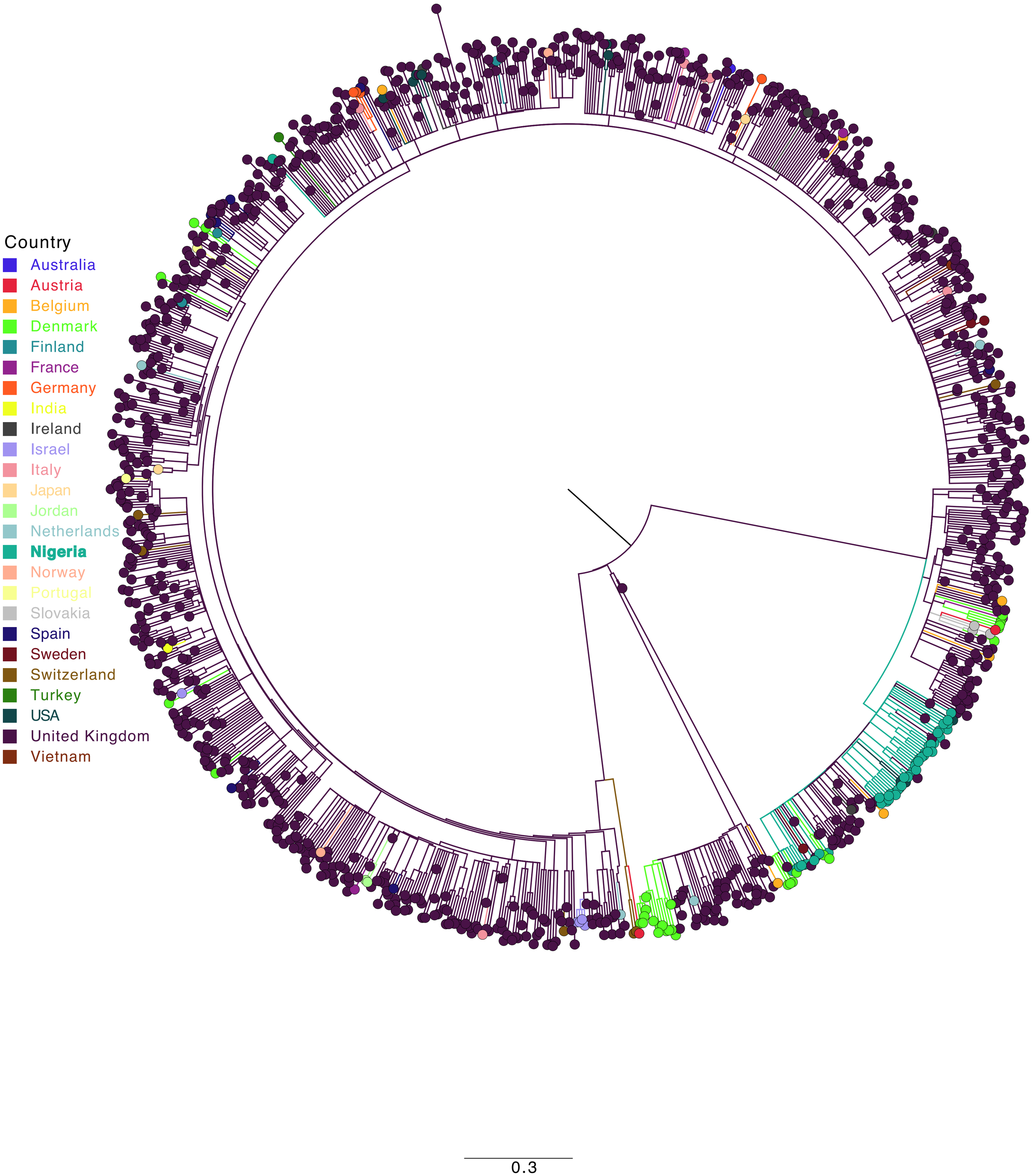
21. Rambaut A, Holmes EC, O'Toole A, Hill V, McCrone JT, Ruis C, et al. A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. Nat Microbiol. 2020.
22. Lauring AS, Hodcroft EB. Genetic Variants of SARS-CoV-2-What Do They Mean? JAMA. 2021;325(6):529-31.
23. Grubaugh ND, Hodcroft EB, Fauver JR, Phelan AL, Cevik M. Public health actions to control new SARS-CoV-2 variants. Cell. 2021;184(5):1127-32.
24. Davies NG, Barnard RC, Jarvis CI, Kucharski AJ, Munday J, Pearson CAB, et al. Estimated transmissibility and severity of novel SARS-CoV-2 Variant of Concern 202012/01 in England. medRxiv. 2020:2020.12.24.20248822.
25. Edara VV, Norwood C, Floyd K, Lai L, Davis-Gardner ME, Hudson WH, et al. Reduced binding and neutralization of infection- and vaccine-induced antibodies to the B.1.351 (South African) SARS-CoV-2 variant. bioRxiv. 2021.
26. Xie X, Liu Y, Liu J, Zhang X, Zou J, Fontes-Garfias CR, et al. Neutralization of SARS-CoV-2 spike 69/70 deletion, E484K and N501Y variants by BNT162b2 vaccine-elicited sera. Nat Med. 2021.
27. Wang L, Zhou T, Zhang Y, Yang ES, Schramm CA, Shi W, et al. Antibodies with potent and broad neutralizing activity against antigenically diverse and highly transmissible SARS-CoV-2 variants. bioRxiv. 2021.
28. Administration USFaD. FDA authorizes revisions to fact sheets to address SARS-CoV-2 variants for monoclonal antibody products under emergency use authorization 2021 [Available from: https://www.fda.gov/drugs/drug-safety-and-availability/fda-authorizes-revisions-fact-sheets-address-sars-cov-2-variants-monoclonal-antibody-products-under.
29. Lemey P, Hong SL, Hill V, Baele G, Poletto C, Colizza V, et al. Accommodating individual travel history and unsampled diversity in Bayesian phylogeographic inference of SARS-CoV-2. Nat Commun. 2020;11(1):5110.
30. Lu L, Lycett S, Ashworth J, Mutapi F, Woolhouse M. What are SARS-CoV-2 genomes from the WHO Africa region member states telling us? BMJ Glob Health. 2021;6(1).
31. Inzaule SC, Tessema SK, Kebede Y, Ogwell Ouma AE, Nkengasong JN. Genomic-informed pathogen surveillance in Africa: opportunities and challenges. Lancet Infect Dis. 2021.
32. (CDC) CfDCaP. CDC 2019-Novel Coronavirus (2019-nCoV) Real Time RT-PCR Diagnostic Panel. Diseases DoV; 2020 3/30/2020. Report No.: CDC-006-00019, Revision 03.
33. Quick J, Grubaugh ND, Pullan ST, Claro IM, Smith AD, Gangavarapu K, et al. Multiplex PCR method for MinION and Illumina sequencing of Zika and other virus genomes directly from clinical samples. Nat Protoc. 2017;12(6):1261-76.
34. Quick J. nCoV-2019 sequencing protocol v2 (GunIt). Apr 9, 2020. Available from: http://dx.doi.org/10.17504/protocols.io.bdp7i5rn.
35. Elbe S, Buckland-Merrett G. Data, disease and diplomacy: GISAID's innovative contribution to global health. Glob Chall. 2017;1(1):33-46.
36. Shu Y, McCauley J. GISAID: Global initiative on sharing all influenza data - from vision to reality. Euro Surveill. 2017;22(13).
37. Nigeria Cenre for Disease Control. NCDC Coronavirus COVID-19 Microsite [Available from: https://covid19.ncdc.gov.ng/.
38. O'Toole Á, Hill V. Global Report of B.1.525 lineage2021. Available from: https://cov-lineages.org/global_report_B.1.525.html.
39. Xie X, Zou J, Fontes-Garfias CR, Xia H, Swanson KA, Cutler M, et al. Neutralization of N501Y mutant SARS-CoV-2 by BNT162b2 vaccine-elicited sera. bioRxiv. 2021:2021.01.07.425740.
40. Galloway SE, Paul P, MacCannell DR, Johansson MA, Brooks JT, MacNeil A, et al. Emergence of SARS-CoV-2 B.1.1.7 Lineage — United States, December 29, 2020–January 12, 2021. Morbidity and Mortality Weekly Report 2021;70(3):95-9.
41. Public Health England. Investigation of novel SARS-CoV-2 variant: Variant of Concern 202012/01, Technical Briefing 3 London, United Kingdom 2021 [Available from:

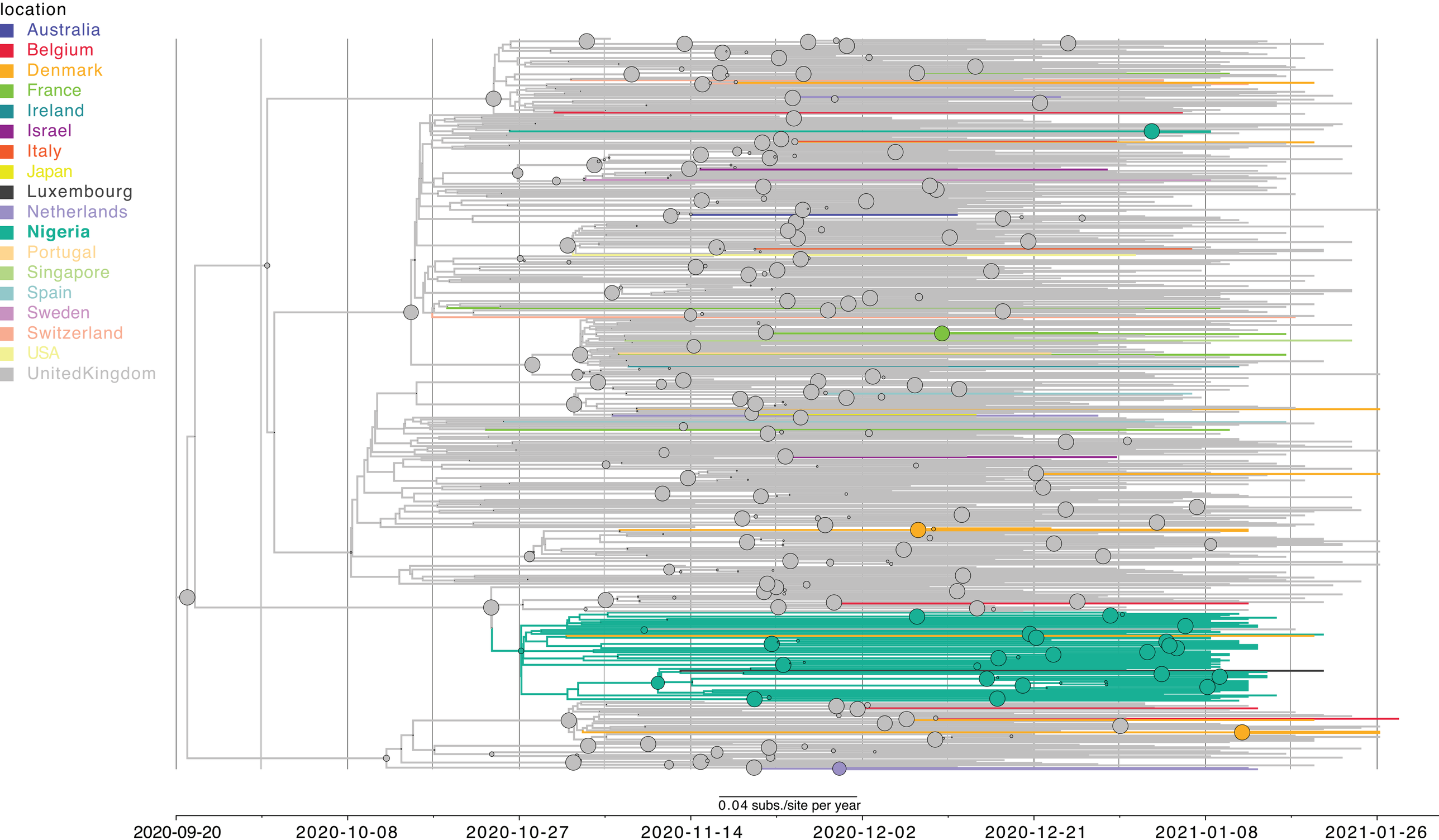
https://assets.publishing.service.gov.uk/government/uploads/system/uploads/attachment_data/file/959360/Variant_of_Concern_VOC_202012_01_Technical_Briefing_3.pdf.
42. Wu K, Werner AP, Moliva JI, Koch M, Choi A, Stewart-Jones GBE, et al. mRNA-1273 vaccine induces neutralizing antibodies against spike mutants from global SARS-CoV-2 variants. bioRxiv. 2021:2021.01.25.427948.
43. Wang Z, Schmidt F, Weisblum Y, Muecksch F, Barnes CO, Finkin S, et al. mRNA vaccine-elicited antibodies to SARS-CoV-2 and circulating variants. bioRxiv. 2021:2021.01.15.426911.
44. Wibmer CK, Ayres F, Hermanus T, Madzivhandila M, Kgagudi P, Oosthuysen B, et al. SARS-CoV-2 501Y.V2 escapes neutralization by South African COVID-19 donor plasma. bioRxiv. 2021:2021.01.18.427166.
45. Greaney AJ, Loes AN, Crawford KHD, Starr TN, Malone KD, Chu HY, et al. Comprehensive mapping of mutations to the SARS-CoV-2 receptor-binding domain that affect recognition by polyclonal human serum antibodies. bioRxiv. 2021:2020.12.31.425021.
46. Weisblum Y, Schmidt F, Zhang F, DaSilva J, Poston D, Lorenzi JC, et al. Escape from neutralizing antibodies by SARS-CoV-2 spike protein variants. Elife. 2020;9.
47. Zhou D, Dejnirattisai W, Supasa P, Liu C, Mentzer AJ, Ginn HM, et al. Evidence of escape of SARS-CoV-2 variant B.1.351 from natural and vaccine induced sera. Cell [Internet]. 2021. Available from: https://doi.org/10.1016/j.cell.2021.02.037.

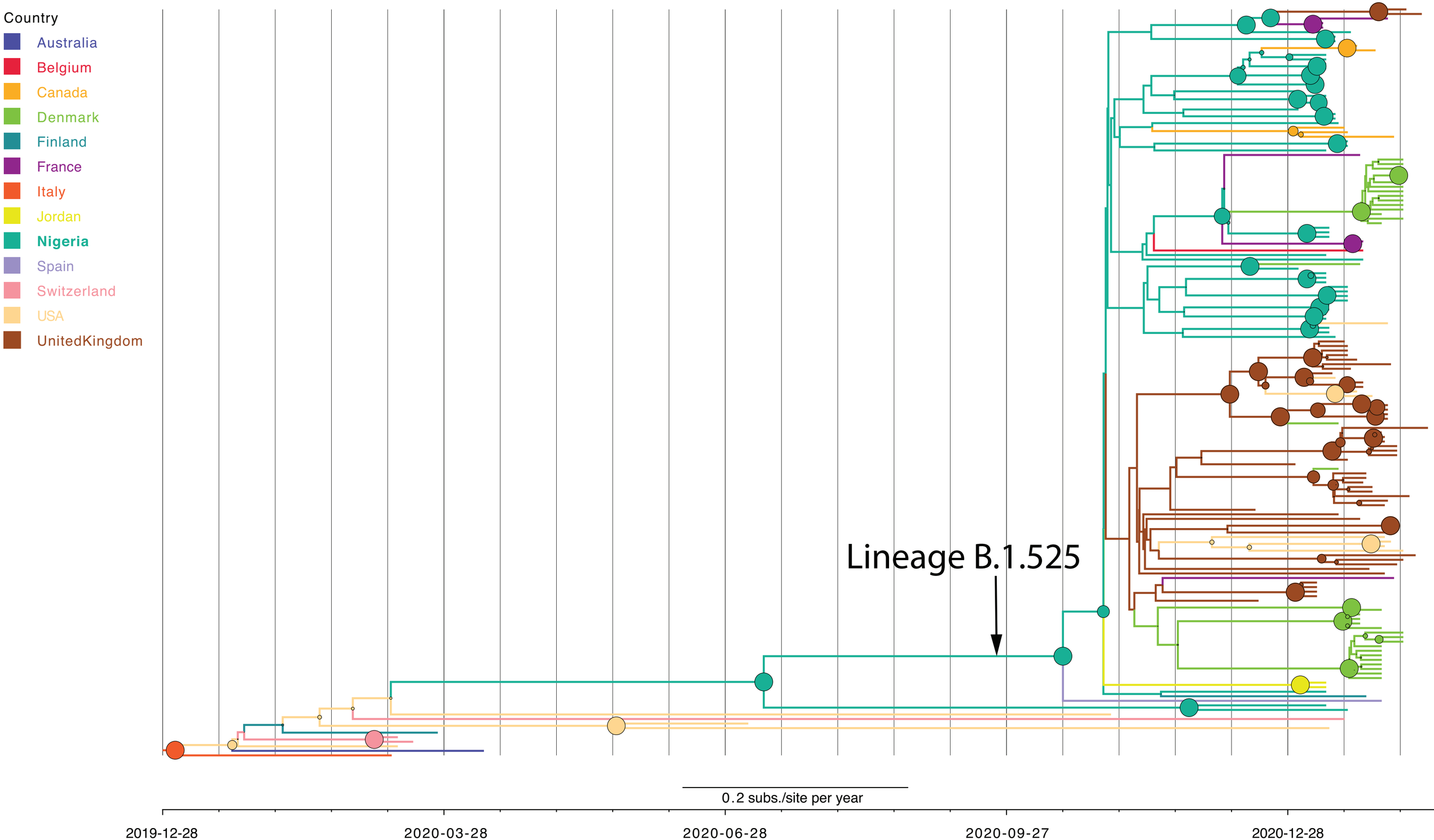

a)
Lineage A
97.5/99
93.2/92
100/100
89.6/88
76.4/94
100/100
99.7/100
84.9/95
85.5/95
Lineage B.1.525
85/98
94.4/70
100/100
97.8/98
82.9/63
91/49
79.4/49
90.2/99
90.1/98
95.3/99
85.9/96
Lineage B.1.1.7
85.7/96
93.2/99
74.3/97
97.8/93
85/86
86.3/99
0.06 subs./site
b)
Frequency (%)
60
40
20
0
A
B.1
B.1.1.10
B.1.1.7
B.1.1.74
B.1.525
lineage

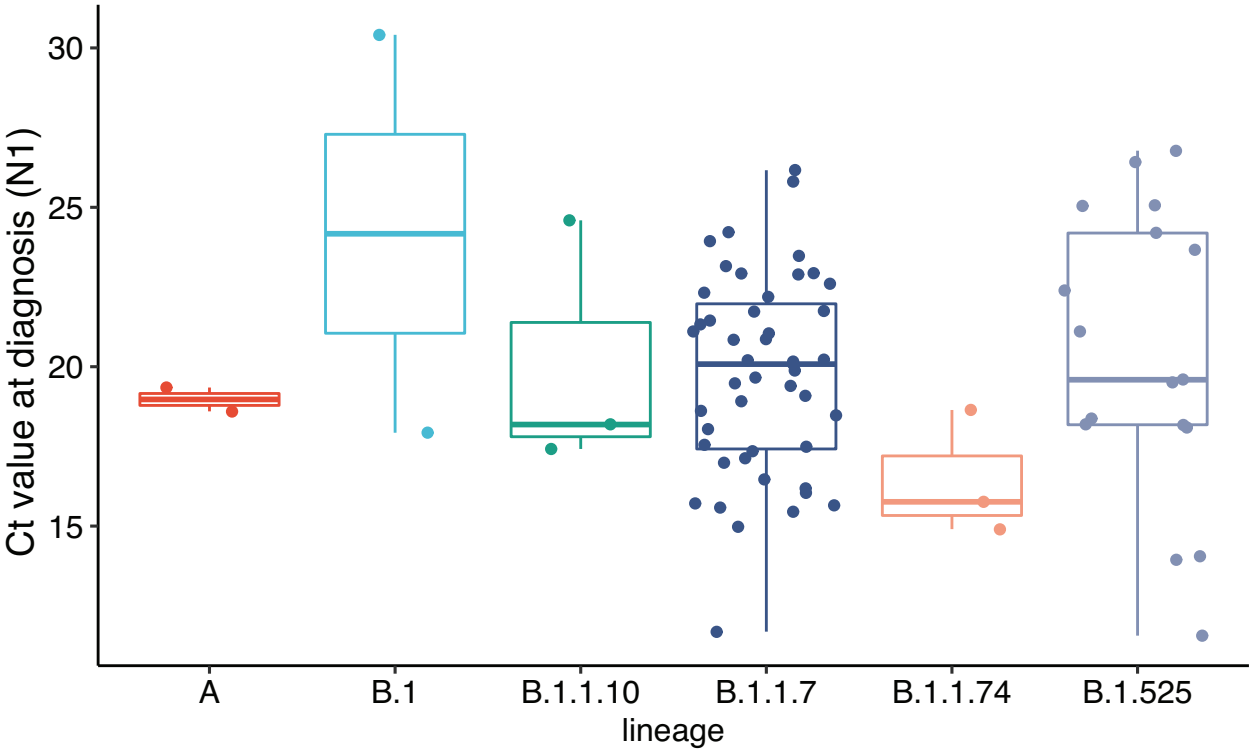

division
Abuja
Akwa-Ibom
Benue
Borno
Delta
Ebonyi
Edo
Ekiti
Enugu
Federal Capital Territory
Kaduna
Kano
Kogi
Kwara
Lagos
Nasarawa
Niger
Ogun
Ondo
Osun
UI
Rivers
Zamfara
nan
UI - Sequences obtained in this study
Lineage B.1.525
Lineage B.1.1.7
0.2
Time Scale
Nov. 2019
Jan. 2021

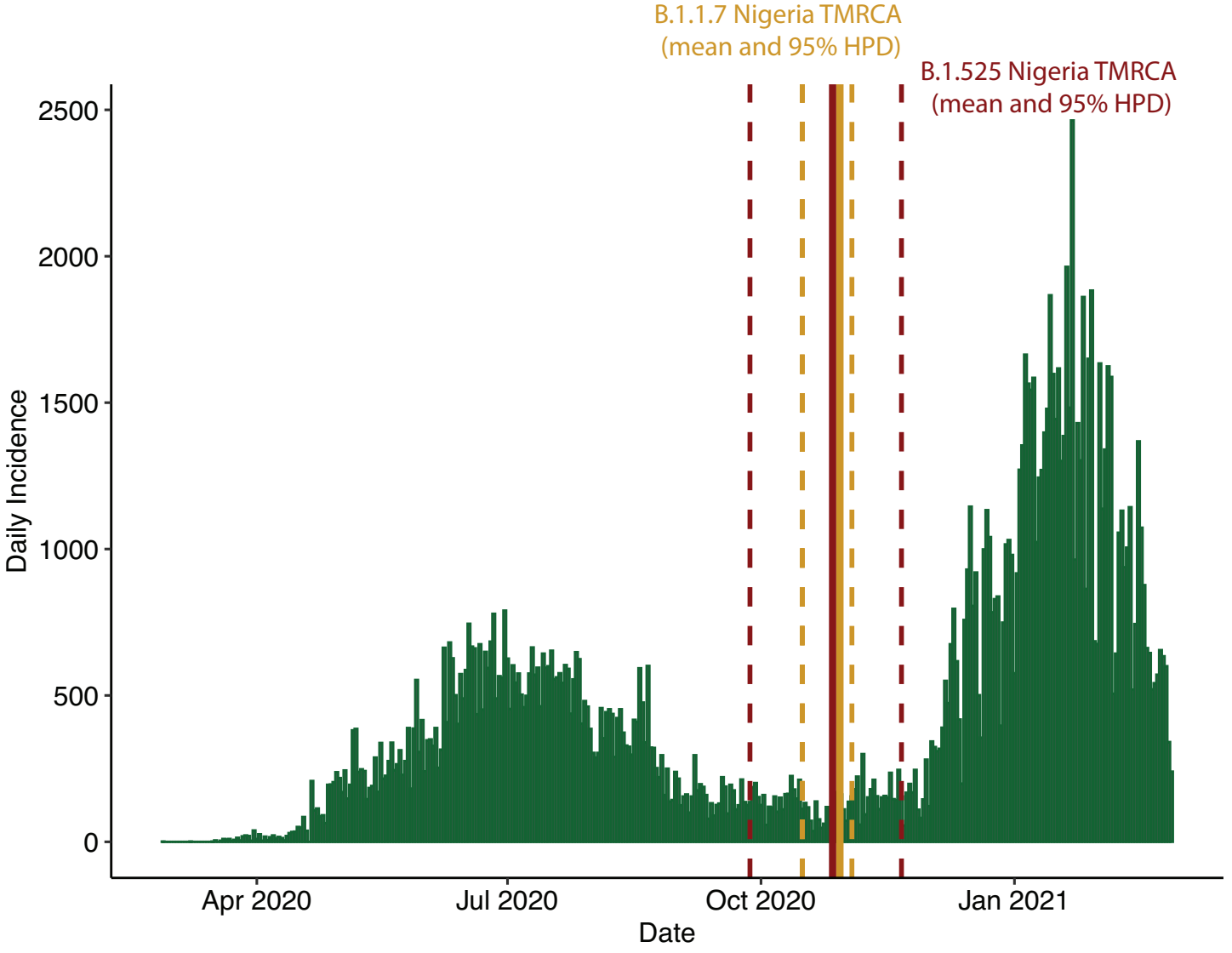

country
Algeria
Andorra
Argentina
Australia
Austria
Bangladesh
Belgium
Brazil
Canada
Chile
China
Colombia
Costa Rica
Czech Republic
DRC
Denmark
Ecuador
Egypt
Finland
France
Gambia
Germany
Greece
Hong Kong
Iceland
India
Indonesia
Iran
Ireland
Israel
Italy
Japan
Jordan
Kazakhstan
Kenya
Latvia
Lithuania
Luxembourg
Madagascar
Malaysia
Malta
Mexico
Myanmar
Netherlands
New Zealand
Nigeria
Norway
Peru
Poland
Portugal
RC
Romania
Russia
Rwanda
Saudi Arabia
Senegal
Serbia
Singapore
Slovakia
South Africa
South Korea
Spain
Sweden
Switzerland
Taiwan
Thailand
Turkey
USA
UAE
UK
Vietnam
Zimbabwe
B.1.1.7
B.1.525
0.3


Country
Australia
Austria
Belgium
Denmark
Finland
France
Germany
India
Ireland
Israel
Italy
Japan
Jordan
Netherlands
Nigeria
Norway
Portugal
Slovakia
Spain
Sweden
Switzerland
Turkey
USA
United Kingdom
Vietnam
0.3


location
Australia
Belgium
Denmark
France
Ireland
Israel
Italy
Japan
Luxembourg
Netherlands
Nigeria
Portugal
Singapore
Spain
Sweden
Switzerland
USA
UnitedKingdom
0.04 subs./site per year
2020-09-20
2020-10-08
2020-10-27
2020-11-14
2020-12-02
2020-12-21
2021-01-08
2021-01-26


Country
Australia
Belgium
Canada
Denmark
Finland
France
Italy
Jordan
Nigeria
Spain
Switzerland
USA
UnitedKingdom
Lineage B.1.525
0.2 subs./site per year
2019-12-28
2020-03-28
2020-06-28
2020-09-27
2020-12-28


Ct value at diagnosis (N1)
30
25
20
15
A
B.1
B.1.1.10
B.1.1.7
B.1.1.74
B.1.525
lineage


B.1.1.7 Nigeria TMRCA
(mean and 95% HPD)
B.1.525 Nigeria TMRCA
(mean and 95% HPD)
Daily Incidence
2500
2000
1500
1000
500
0
Apr 2020
Jul 2020
Oct 2020
Jan 2021
Date