











# Multiple expansions of globally uncommon SARS-CoV-2 lineages in Nigeria

Egon A. Ozer <sup>1,2,11</sup>, Lacy M. Simons <sup>1,2,11</sup>, Olubusuyi M. Adewumi <sup>3,4,11</sup>, Adeola A. Fowotade<sup>4,5</sup>, Ewean C. Omoruyi<sup>5</sup>, Johnson A. Adeniji<sup>3,4</sup>, Oluseyi A. Olayinka<sup>3,4</sup>, Taylor J. Dean<sup>1,2</sup>, Janet Zayas<sup>6</sup>, Pavan P. Bhimalli<sup>6</sup>, Michelle K. Ash<sup>6</sup>, Almoustapha I. Maiga<sup>7</sup>, Anou M. Somboro <sup>7,8</sup>, Mamoudou Maiga<sup>7,9</sup>, Adam Godzik<sup>10</sup>, Jeffrey R. Schneider<sup>6</sup>, João I. Mamede <sup>6</sup>, Babafemi O. Taiwo<sup>1</sup>, Judd F. Hultquist <sup>1,2</sup>  & Ramon Lorenzo-Redondo <sup>1,2</sup> 

Disparities in SARS-CoV-2 genomic surveillance have limited our understanding of the viral population dynamics and may delay identification of globally important variants. Despite being the most populated country in Africa, Nigeria has remained critically under sampled. Here, we report sequences from 378 SARS-CoV-2 isolates collected in Oyo State, Nigeria between July 2020 and August 2021. In early 2021, most isolates belonged to the Alpha “variant of concern” (VOC) or the Eta lineage. Eta outcompeted Alpha in Nigeria and across West Africa, persisting in the region even after expansion of an otherwise rare Delta sub-lineage. Spike protein from the Eta variant conferred increased infectivity and decreased neutralization by convalescent sera in vitro. Phylodynamic reconstructions suggest that Eta originated in West Africa before spreading globally and represented a VOC in early 2021. These results demonstrate a distinct distribution of SARS-CoV-2 lineages in Nigeria, and emphasize the need for improved genomic surveillance worldwide.

<sup>1</sup> Division of Infectious Diseases, Department of Medicine, Northwestern University Feinberg School of Medicine, Chicago, IL, USA. <sup>2</sup> Center for Pathogen Genomics and Microbial Evolution, Northwestern University Hayek Institute for Global Health, Chicago, IL, USA. <sup>3</sup> Department of Virology, College of Medicine, University of Ibadan, Ibadan, Nigeria. <sup>4</sup> Infectious Disease Institute, College of Medicine, University of Ibadan, Ibadan, Nigeria. <sup>5</sup> Biorepository and Clinical Virology Laboratory, College of Medicine, University College Hospital, University of Ibadan, Ibadan, Nigeria. <sup>6</sup> Department of Microbial Pathogens and Immunity, Rush University Medical Center, Chicago, IL, USA. <sup>7</sup> University Clinical Research Center (UCRC), University of Sciences, Techniques et Technologies of Bamako (USTTB), Bamako, Mali. <sup>8</sup> School of Laboratory Medicine and Medical Sciences, University of KwaZulu-Natal, Durban, South Africa. <sup>9</sup> Biomedical Engineering and Preventive Medicine Department, Northwestern University, Evanston, IL, USA. <sup>10</sup> Biosciences Division, University of California Riverside School of Medicine, Riverside, CA, USA. <sup>11</sup> These authors contributed equally: Egon A. Ozer, Lacy M. Simons, Olubusuyi M. Adewumi. email: [judd.hultquist@northwestern.edu](mailto:judd.hultquist@northwestern.edu); [ramon.lorenzo@northwestern.edu](mailto:ramon.lorenzo@northwestern.edu)

Almost two years 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<sup>1–3</sup>. Despite the relatively low mutation rate of this virus, its high prevalence in the human population globally has allowed it to diversify quickly<sup>4,5</sup>. 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<sup>6–9</sup>. For example, during the spring and summer of 2020, a mutation in the viral Spike protein, D614G, was identified in association with higher viral loads in patient upper airways<sup>10,11</sup>. The rapid expansion of this variant in communities across the world coupled with virological assays *in vitro* and *in vivo*, suggested that this variant is more transmissible than earlier lineages<sup>12–15</sup>. Indeed, this mutation is nearly fixed in the global SARS-CoV-2 population today<sup>11</sup>.

Continued surveillance efforts by public health entities such as the World Health Organization (WHO) and Centers for Disease Control (CDC) have designated ‘variants of concern (VOC)’ associated with increased transmissibility, increased disease severity, decreased susceptibility to therapeutic agents, and/or decreased susceptibility to antibody neutralization<sup>16</sup>. Lineages classified as ‘variants of interest (VOI)’ are those with genetic markers predicted to affect transmission, immune escape, therapeutic resistance, and/or diagnostic evasion<sup>17</sup>. These designations reflect concerns that more transmissible variants could dampen the effectiveness of current public health practices designed to limit the spread of the virus<sup>18–20</sup>. Some mutations may additionally impact considerations for clinical care by either altering severity of the disease or dampening the effectiveness of therapeutic monoclonal antibodies or other antiviral treatments<sup>21</sup>. Finally, changes to the viral Spike protein may render current vaccine formulations less efficacious and/or confer the increased capacity for re-infection<sup>22–24</sup>.

As the COVID-19 pandemic has progressed and evolved, designations of lineages as VOCs or VOIs has changed and has often differed between public health entities. Throughout early- and mid-2021, there were four VOCs named using the dynamic Pangolin nomenclature for SARS-CoV-2 (cov-lineages.org) and the WHO variant naming scheme: the B.1.1.7 lineage (Alpha variant) originally identified in the United Kingdom, the B.1.351 (Beta variant) originally identified in South Africa, the P.1 (Gamma variant) originally identified in Brazil, and the most recent B.1.617.2 and related AY lineages (Delta variant) initially identified in India<sup>25–30</sup>. As of late September 2021, as Delta has become the predominant variant worldwide, the CDC has updated its VOC classification to include only Delta<sup>31</sup>. Previously VOIs included the B.1.525 (Eta), B.1.526 (Iota), C.37 (Lambda), and B.1.621 (Mu) lineages, however both the WHO and CDC have recently removed all lineages from the VOI classification and have instead designated these and others as ‘variants under monitoring (VUM)’ or ‘variants being monitored (VBM)’, respectively<sup>31</sup>. Addressing the threats posed by variants of concern and other variants with enhanced epidemic or pandemic potential requires continued surveillance of SARS-CoV-2 genetic diversity worldwide and the ongoing assessment of the impact of novel mutations or combinations of mutations on viral fitness, transmission, antibody neutralization, therapeutics, and pathogenesis.

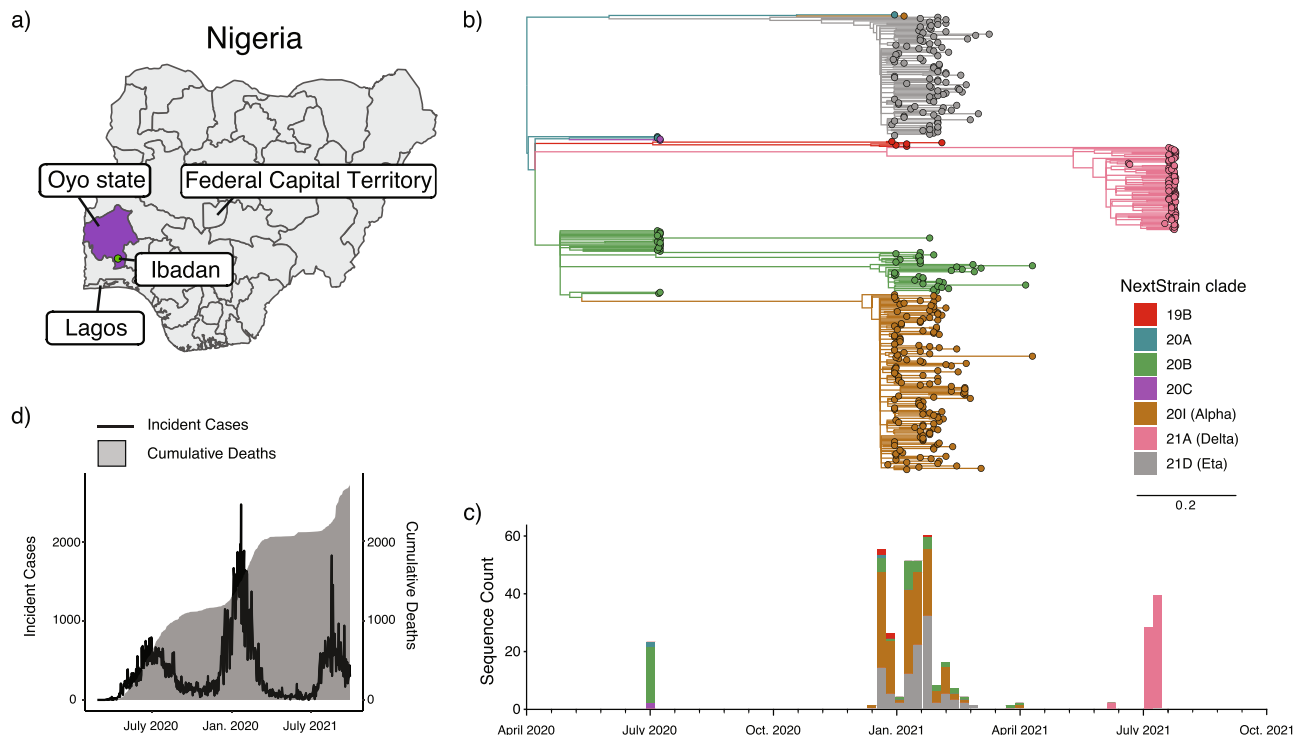
While many countries have developed extensive genetic surveillance and reporting systems, several regions across the globe remain critically undersampled<sup>32</sup>. For example, despite remarkable efforts by multiple groups and institutions<sup>33</sup>, several countries in Africa have reported only a handful of sequences relative to their cumulative case counts, if any<sup>34,35</sup>. Nigeria, the most populated country in Africa and the seventh most populated country

worldwide, has only a small number of reported SARS-CoV-2 sequences available in public repositories (currently just over 1400). Through the beginning of 2021, the majority of available sequences had been generated through the African Centre of Excellence for Genomics of Infectious Disease (ACEGID), which serves a number of African countries ([www.acegid.org](http://www.acegid.org)), but more recent data has been limited. Given Nigeria’s status as an epicenter of commerce and travel in Africa, undetected expansion of a more infectious, virulent or immune-resistant variant could have major repercussions for the wider region or the entire continent. A more consistent and higher volume of sample collection and sequencing in Nigeria is required to strengthen the public health value of COVID-19 surveillance efforts.

To better understand the SARS-CoV-2 population structure in Nigeria in 2021, to add more sequences from this region to public repositories, and to perform longitudinal analyses of viral variation concurrent with the highest incidence period in the country, we performed viral whole-genome sequencing on 378 SARS-CoV-2 isolates collected from COVID-19 patients in Oyo state, Nigeria between July of 2020 and August of 2021. Oyo state’s capital city, Ibadan, is the third most populous city in Nigeria situated ~130 km away from Lagos International Airport. Oyo state is the ideal location to study both imported and locally circulating SARS-CoV-2 lineages due to its proximity to the Republic of Benin and an international airport, and its blended composition of urban and rural areas. In addition, Oyo is home to the University of Ibadan College of Medicine that provides health services for the entire region, casting a wide network for sample collection across Nigeria. Phylogenetic analysis revealed two primary lineages circulating in the region during early 2021, the B.1.1.7 or ‘Alpha’ VOC, and the B.1.525 or ‘Eta’ lineage, previously designated as a VOI. The Eta lineage outcompeted Alpha and even persisted in the region after the introduction of the Delta VOC in April of 2021. Using a pseudotyped virus system, we found that mutations in the Spike protein of Eta isolates enhanced viral entry *ex vivo*. Furthermore, these mutations decreased the effectiveness of antibodies from previously infected individuals to neutralize the virus, though antibodies from individuals immunized with mRNA vaccines were still effective. Interestingly, while Delta variant became dominant in Nigeria similarly to what can be observed globally, we observe that a majority of cases in Nigeria are due to a rare Delta lineage, AY.36, which constitutes over 80% of Delta cases in Nigeria, but less 0.5% of Delta cases globally. Together, these data add to SARS-CoV-2 genomic surveillance in Nigeria, identify two lineages of note that dominated the epidemic during the period of highest incidence in the region, suggest that the Eta lineage may share features with previous and current variants of concern, and identify the recent dominance of very uncommon Delta lineage. Improved and ongoing genomic surveillance in under-sampled regions across the globe will be necessary for the effective and accurate understanding of potential health risks and for their effective communication to communities worldwide.

## Results

**Specimen characteristics.** We obtained 421 specimens from individuals with COVID-19 in Oyo state, Nigeria (Fig. 1a) collected between 12 July 2020 and 3 August 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 Biorpository and Clinical Virology Laboratory at the University College Hospital, College of Medicine at the University of Ibadan. The qRT-PCR cycle threshold (Ct) values of these specimens



**Fig. 1** Phylogenetic analysis of SARS-CoV-2 isolates in Oyo state. **a** Geographical location of Oyo state, Nigeria is indicated (purple), where the sampling for this study was performed. **b** ML phylogenetic tree of 378 SARS-CoV-2 specimen genomes in Oyo state collected between 12 July 2020 and 3 August 2021. Clades following Nexclade nomenclature are indicated and colored. Scale indicates the number of substitutions per site and time. **c** Distribution of the different clades per time found in the Oyo dataset reported here. **d** Evolution of daily COVID-19 incidence and cumulative daily COVID-19-related deaths.

obtained in the clinical laboratory ranged from 13.46 to 42.87 for the N target (either DaAn Gene or BGI detection kits) and 14.23 to 41.56 for the ORF1ab target (DaAn Gene kit). The average age of the patients was 42 years (range 4–86) and the sex distribution was 44% female, 37% male, and 18% unknown. RNA was extracted from each sample and the presence of viral RNA was confirmed by one-step qRT-PCR (CDC assay, RNaseP, and N1 primer set)<sup>36</sup>. RNA samples of sufficient quality (RNaseP control, Ct value <35) and with sufficient copies of the viral genome for 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 subjected to deep sequencing on the Illumina platform<sup>37,38</sup>. The minimum threshold for base calling was 10 reads with 90% coverage required across the genome to report the whole-genome sequence. Of the 421 specimens, 43 failed to yield a satisfactory 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 378 complete SARS-CoV-2 genomes were deposited in the public GISAID database (Supplementary Data 1) and subjected to phylogenetic analysis<sup>39,40</sup>.

**The lineage of concern B.1.1.7 and the lineage of interest B.1.525 dominated the epidemic in Oyo state until the arrival of the Delta lineages.** We first performed phylogenetic reconstruction using maximum likelihood (ML) phylogenetic analysis of the 378 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 (Fig. 1b). Between January 2021 and August 2021 most of the specimens from Oyo state (89%) belonged to one of three main clusters with very strong support at their base nodes (support: >97% aLRT and

>97% UFboot). We used the Pango classification scheme (PangoLearn version 9/28/2021) to identify the lineages of the major clusters. The most abundant lineage from January to April 2021 was B.1.1.7, or Alpha (147 out of 287 sequences). This lineage is associated with the Spike N501Y mutation and was first identified in the U.K. in the fall of 2020<sup>28</sup>. The second most common lineage during this early part of 2021 was B.1.525, or Eta (100 of 287 sequences), which expanded over time despite being in direct competition with Alpha (Fig. 1c). 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 and Spike 144del (cov-lineages.org). All sequences obtained between June and August 2021 belonged to the Delta variant lineages B.1.617.2 (2 sequences), AY.36 (60 sequences), and AY.4 (7 sequences). Interestingly, the most prevalent Delta lineage in this set of isolates, AY.36, is currently at a very low prevalence worldwide, represented by just under 0.5% of the Delta sequences as of September 2021 (outbreak.info). The remaining sequences all belonged to other, previously common B.1 lineages, except for three sequences from lineage A and 2 from A.27. Viruses from lineage A and derivatives spread throughout the world early in the pandemic, but are now relatively rare globally, accounting for <0.5% of isolates sequenced and deposited in August 2021 (nextstrain.org).

To better determine if these sequences in Oyo state were representative of other parts of Nigeria, we downloaded all available sequences sourced from Nigeria in the GISAID database (<https://www.gisaid.org/>) as of 3 August 2021 (856 sequences, excluding those reported in this study). This included sequence data from Lagos state, Osun state, Abuja, and other sequences from Oyo state collected outside this study. 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 other Nigerian states. This analysis confirmed a distribution of the viral populations throughout the country similar to what was observed in Oyo state, but with a higher proportion of the Eta lineage. Of all sequences collected between 1 January 2021 and 30 April 2021 ( $n = 875$ ), we observed that 46% of the sequences belonged to the Eta lineage while 32% belonged to the Alpha lineage (Supplementary Fig. 1). After 30 April 2021 ( $n = 91$ ), 89% of the sequences belonged to Delta lineages while only 2% of the sequences belonged to Eta lineages; no sequences from Alpha lineages were isolated. Among the Delta lineages, we confirmed the high prevalence of AY.36, indicating that this was not a local observation from Oyo state. Overall, these results suggest that the distribution of lineages observed in this study was broadly reflective of the contemporaneous distribution of lineages in Nigeria, and confirms the prevalence of the Eta and Alpha lineages coincident with the period of highest overall COVID-19 case counts in the country<sup>41</sup> (Fig. 1d). These lineages were subsequently replaced by the Delta lineages coincident with the third spike in cases observed after July 2021.

**Introduction and characteristics of the Alpha, Eta, and Delta lineages in Nigeria.** To better understand these data in the context of the global epidemic, we subsampled global sequences from GISAID collected through 3 August 2021 from these three main lineages found in Nigeria and repeated these analyses. First, we analyzed the Alpha sequences sampled in Nigeria in the context of globally distributed Alpha sequences and performed a ML phylogenetic reconstruction (Supplementary Fig. 2). These analyses showed that most Alpha lineage isolates in Nigeria were closely related and likely had a monophyletic origin with subsequent diversification (aLRT = 90.1; Supplementary Fig. 2). This sub-lineage included a prominent sub-cluster with an additional, common mutation (aLRT = 85.1; Supplementary Fig. 2). This result suggests either one successful introduction of Alpha in Nigeria, or multiple, closely linked introductions compatible with the observed founder effect. To analyze this further, we performed Bayesian inference of all Alpha sequences available from Nigeria alongside a smaller set of subsampled Alpha sequences based on time and sequence diversity from across the globe (Supplementary Fig. 3). These analyses confirmed strong statistical support for a monophyletic origin of almost all of the Alpha sequences from this study with the most likely origin being from the United States. Using this approach, we also estimated the time to the most recent common ancestor (TMRCA) for the Alpha sub-lineage in Nigeria to be around 22 October 2020 [95% Highest Posterior Density (HPD) interval: 12 October–30 October 2020].

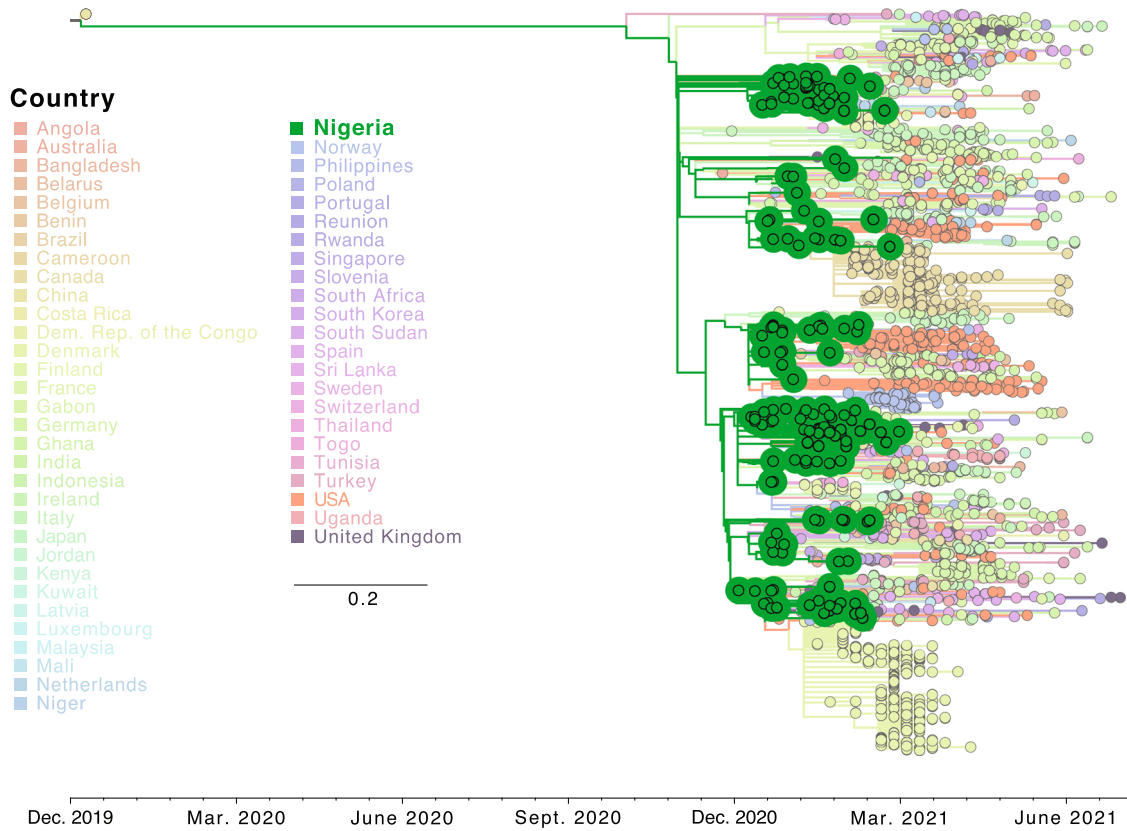
The second most prevalent lineage observed in our dataset, Eta, was not as highly prevalent on a global scale as Alpha, but was previously considered a VOI and has been defined as a lineage of international significance<sup>42</sup>. When first detected in February of 2021 as part of our group's ongoing genomic surveillance efforts, only 159 similar sequences were available in GISAID, a plurality (25%) of which originated in Nigeria<sup>42</sup>. Our continuing sampling during the first months of 2021 provided sufficient sequences for lineage designation and confirmed the high prevalence of this lineage in Nigeria compared with many other places around the globe. Lower prevalence enabled us to perform a ML analysis of our sequences with all Eta lineage sequences available in GISAID as of 3 August 2021 ( $n = 2329$ , not including the sequences from this study). We observed strong statistical support for a monophyletic clade of the Eta lineage in Nigeria close to the

origin point, suggestive of a single introduction followed by rapid expansion soon after the initial appearance of this lineage (Fig. 2). Using a Bayesian approach, we next analyzed all Eta isolates from Oyo state relative to a subsample based on time and diversity of all Eta sequences in GISAID (Supplementary Fig. 4). The estimated TMRCA for Eta in Nigeria was 16 October 2020 [95% HPD interval: 21 September–8 November 2020], similar to the TMRCA estimated for the Alpha lineage. Together, these analyses suggest that the Eta lineage may have originated from or expanded principally in Nigeria before spreading in West Africa and globally.

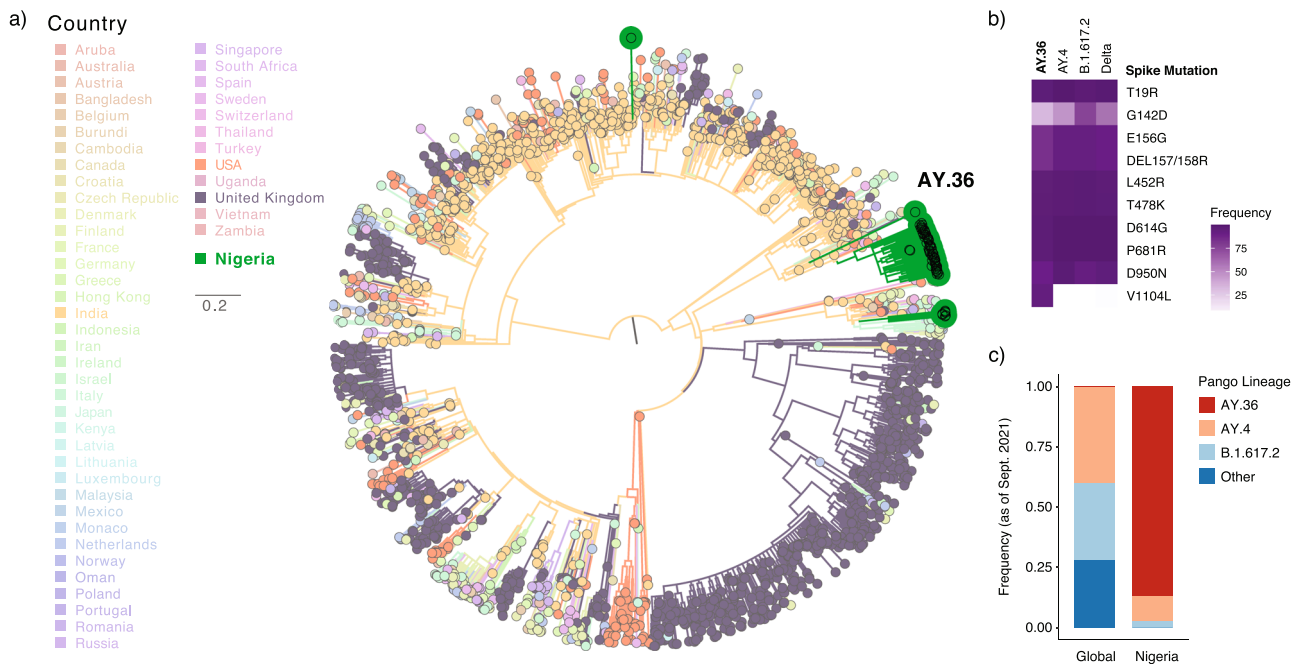
Nearly all isolates analyzed in this study after June of 2021 belonged to one of the Delta lineages. The Delta variant has become predominant in most countries across the globe, including Nigeria. We performed a ML analysis of the Delta sequences from this study with a temporal subsample of global Delta sequences and found that a majority of sequences appeared to be monophyletic with very strong statistical support (Fig. 3). Indeed, 84% of Delta isolates in this study were classified as lineage AY.36, an uncommon Delta lineage that as of 3 August 2021 constituted <0.5% of global Delta sequences. Due to the rarity of AY.36, this analysis strongly suggests that most of the Delta variant sequences observed in Nigeria derived from a single successful introduction of AY.36 and its subsequent expansion in the country (Supplementary Fig. 5). The estimated TMRCA for AY.36 sequences in Nigeria was inferred to be 25 April 2021 [95% HPD interval: 1 April 2021–13 May 2021].

Taken together, these analyses are consistent with a single introduction or multiple, closely linked introductions of both the Alpha and Eta lineages into Nigeria in the fall of 2020. These introductions were followed by the expansion of both populations to become the predominant lineages in the country, with a wider expansion of Eta by the end of the spring of 2021. The epidemic in Nigeria has been subsequently dominated by an uncommon Delta lineage (AY.36) introduced in late April 2021, with a remnant presence of the Eta variant.

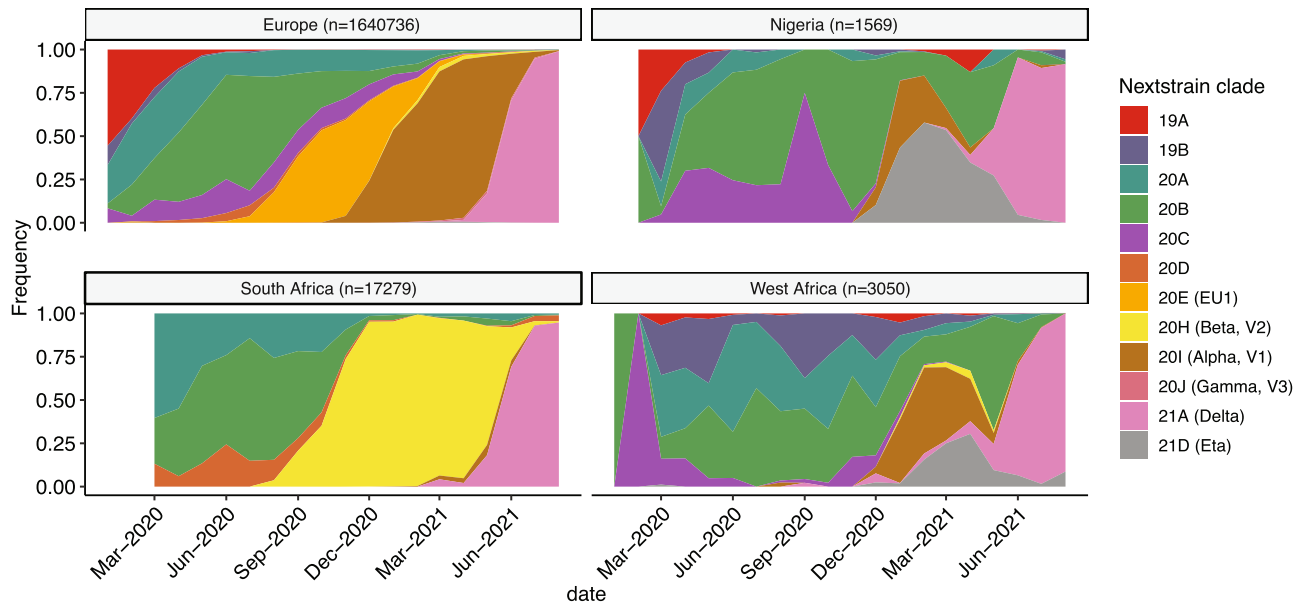
**Comparison of the SARS-CoV-2 population structure in Nigeria to other global regions.** The previous results suggested that the epidemic in Nigeria in early 2021 was defined by the concurrent rise of two competing variants of note, the Alpha VOC and the Eta VOI, followed by the rise of an uncommon Delta variant, AY.36. We next compared the temporal distribution of variants in Nigeria over the course of the pandemic to the rest of West Africa (excluding Nigeria), South Africa (the most widely sampled African country), and Europe (as an external and widely sampled region) (Fig. 4). This analysis shows a temporal and proportional distribution of lineages in Nigeria that is very similar to the rest of West Africa, but distinct from either Europe or South Africa. This is particularly notable in early 2021, where Alpha becomes predominant in Europe, the Beta variant becomes predominant in South Africa, but the pandemic in West Africa is defined by a competition between Alpha and Eta. These data further showed that the Eta variant appeared to outcompete Alpha in both Nigeria and West Africa as a whole before the emergence and introduction of the Delta lineages. Indeed, the Eta variant continues to circulate in West Africa as a minority of isolates in September of 2021 (<https://www.gisaid.org/>), even while Delta has nearly completely replaced other SARS-CoV-2 variants in other regions of the world. Moreover, to further examine the resilience of this lineage, we sampled 50 isolates (from which we were able to amplify 37 viral genomes) collected between December 2020 and August 2021 from a non-neighboring West African country, Mali. During this time period 81% of the sequences from Mali were the Eta variant,



**Fig. 2 Phylogenetic analysis of the entire B.1.525 lineage.** ML phylogenetic temporal reconstruction of full genome sequences from all full genome B.1.525 sequences from this study and from GISAID as of 3 August 2021. Clades corresponding to Nigerian sequences are indicated. Branches and tips are colored by country. Scale indicates number of substitutions per site and time.



**Fig. 3 Analysis of Oyo state Delta sequences compared to global Delta sequences.** **a** ML phylogenetic temporal reconstruction of full genome Delta sequences obtained in this study and temporally sampled Delta global sequences from GISAID as of 3 August 2021. Branches and tips are colored by country. The clade corresponding to the AY.36 lineage is indicated. **b** Spike protein mutation prevalence across all Delta lineages and the 2 most frequent lineages (AY.4 and B.1.617.2) compared to AY.36 lineage. **c** Prevalence of the most frequent Delta lineages globally compared to the distribution in Nigeria where AY.36 lineage was the most prevalent. AY.36 was represented in the global distribution (only 0.2% of global sequences) for comparison purposes. Data for **b** and **c** was obtained from (<https://outbreak.info/>)<sup>84</sup>.



**Fig. 4 Temporal SARS-CoV-2 clade distribution in Nigeria and West Africa.** Monthly clade frequency distribution using Nextclade nomenclature for clades with a frequency >1% in both Nigeria and West Africa compared to the main clades observed in Europe and South Africa. All GISAID sequences deposited up to 30 September 2021 were used to examine the period of Delta dominance. The number of sequences used per geographical location is indicated. Nextclade nomenclature was used due to its clarity to represent major clades.

including 4 out of 4 obtained in July 2021 and 2 out of 4 obtained in August 2021 (Supplementary Fig. 6).

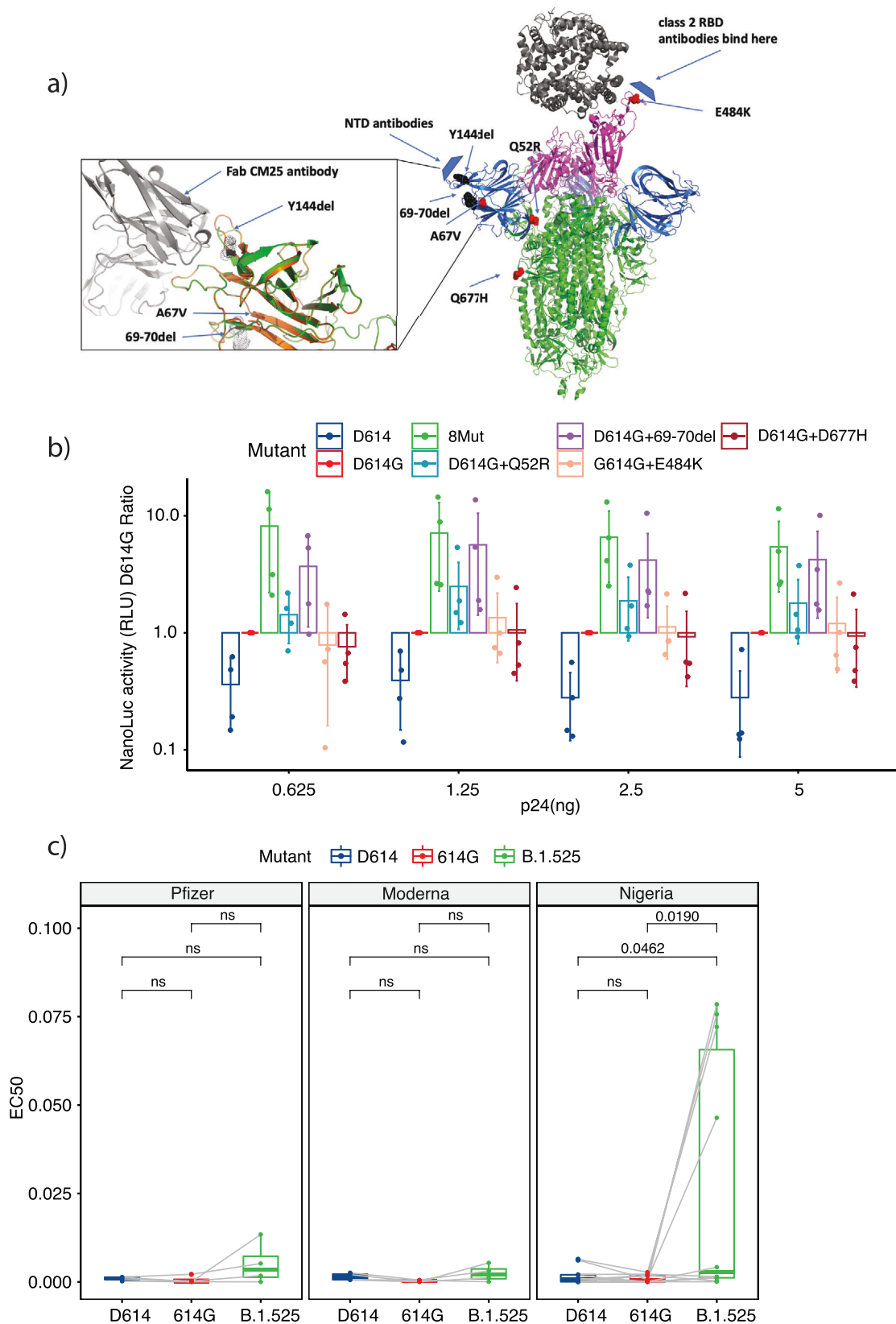
**Limited differences in the Ct values of the diagnostic PCR tests by lineage.** The Alpha and Delta lineages have been previously associated with increased viral loads in the upper airways of patients relative to other SARS-CoV-2 lineages<sup>43</sup>. To assess this in our study, we compared the real-time PCR cycle threshold (Ct) values for the N1 probe measured for each isolate at the time of diagnosis as a proxy for viral load. Using a linear model to control for patient age and sex, we compared the Ct value for all 378 diagnostic specimens batched by lineage (Supplementary Fig. 7). We did not observe any statistically significant differences between the three main variants (Alpha, Eta, and Delta lineages), though both Alpha and Delta had significantly lower Ct values than 20B as previously reported<sup>43</sup>. Although a larger dataset that controls for a higher number of 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 three lineages at the time of patient diagnosis.

**Spike protein mutations in the B.1.525 lineage promote cell entry.** Given that the expansion of Eta lineage in Nigeria occurred alongside the expansion of the Alpha VOC, we sought to characterize the phenotypic impact of the Eta Spike (S) mutations on viral infectivity. Eta lineage isolates are characterized by several nonsynonymous mutations in S at positions Q52R, A67V, E484K, D614G, Q677H, and F888L, as well as by two in-frame deletions at positions 69–70 and 144. Notably, both deletions are also found in the Alpha lineage and have been previously linked with enhanced infectivity and transmission<sup>44</sup>, while the E484K mutation has arisen independently in multiple lineages and has been associated with the potential for immune evasion<sup>45</sup>.

Mapping the Eta lineage-defining mutations onto available Spike protein structures demonstrated an accumulation of changes in the N-terminal domain (NTD) of the protein. To better understand the impact of these mutations, we used

structural modeling to examine the predicted changes in Spike protein structure in the presence and absence of binding to both ACE2 as well as NTD-directed antibody (Fig. 5a). As suggested by studies in other lineages, the E484K mutation in the receptor-binding domain (RBD) modifies a main epitope for class 2 antibodies. In addition, several other mutations in the Eta lineage are immediately adjacent to another antigenic “supersite” located in the NTD<sup>46–48</sup>, independent of the RBD. Deletion of the residue Y144 in particular directly impacts the central ‘N3 loop’ in the supersite while the 69–70 deletion and the A67V mutation further shift this loop. The Q52R mutation is also located in the NTD, but further from the identified antigenic region. Q677H is close to the furin cleavage site, potentially affecting the dynamics of Spike cleavage, a process required for cell entry<sup>49</sup>. Q677H is also seen in minor variants of Alpha and Delta, including a fast-growing AY.34 subvariant, and is currently considered a mutation of interest for increased surveillance. Finally, the F888L mutation is located in the CD1 (connecting domain 1) region of the S2 spike subunit, and there is increasing evidence that mutations in the S2 region may act synergistically with furin region mutations to increase the efficiency of the fusion process.

To test the impact of these mutations on viral entry, we generated a panel of HIV-1 luciferase reporter viruses pseudotyped with SARS-CoV-2 Spike containing the selected mutations. Given that the Spike D614G mutation is nearly fixed on the global scale, this construct was used as a baseline. On this background, each of the other eight mutations in B.1.525 were introduced both individually (with the exception of A67V that was initially not considered a lineage-defining mutation and the 144 deletion/F888L mutation that we were not able to generate) as well as cumulatively to generate an Eta lineage consensus Spike. The ancestral SARS-CoV-2 Spike D614 was also included as a control. Viral input was normalized by the concentration of the HIV p24 protein and used to challenge HeLa cells overexpressing ACE2 in step-dilution MOI in technical duplicates across two independent experiments. Successful delivery of the reporter gene was monitored by luciferase activity in the cell lysate 48 h after inoculation (Fig. 5b and Supplementary Fig. 8a). As previously



reported<sup>10-15</sup>, the D614G mutation significantly enhanced the delivery of the reporter over the original Spike containing D614. Two mutations enhanced reporter activity even further in conjunction with the D614G mutation: the 69-70 deletion (FDR < 0.001) and the Q52R point mutation (FDR = 0.0137). Moreover, the combination of mutations contained in the Eta

consensus Spike demonstrated the highest amount of viral entry relative to each tested mutation indicating an additive effect of these mutations. A western blot of virus-like particles used for cell challenge demonstrated that all generated mutants incorporated Spike protein properly into viral particles (Supplementary Fig. 8b). Together, these results suggest that the Spike mutations in the Eta

**Fig. 5 Analysis of B.1.525 Spike mutations on cellular entry and antibody neutralization.** **a** Overall structure of the SARS-CoV-2 spike protein trimer with N-terminal domains (NTD) in blue, receptor-binding domains (RBD) in magenta, and ACE2 receptor in dark gray, based on the set of PDB coordinates 7a94<sup>85</sup> showing the spike protein bound to one ACE2 molecule. Mutated residues in the B.1.525 spike are shown in chain A as red spheres, deleted residues as black spheres. A model of the B.1.525 NTD in the inset is shown interacting with the Fab C25 antibody (PDB 7m8j)<sup>81</sup>. The unmodified chain on the wild-type NTD is shown in orange. A significant conformational change in the N3 loop results in a significant drop in the (estimated) binding energy to this antibody, from  $-10.8$  kcal/mol to  $-1.3$ . **b** Nanoluc activity measured in relative light units (RLU) ratio between each of the mutants tested and the D614G mutant. D614G was used to calculate the ratios due to its predominance in the population before the appearance of the B.1.1.7 and B.1.525 lineages. Values are shown for each of the dilutions used after p24 concentration normalization. Dots represent the value of each of the replicates performed per mutant ( $n = 4$ ). Bars represent the mean and lines represent the standard deviation of the replicates. **c** Neutralization  $EC_{50}$  comparison between the different Spikes tested in the presence of sera from vaccinated (Pfizer [ $n = 4$ ] or Moderna [ $n = 4$ ] groups) or naturally-infected (Nigeria group [ $n = 9$ ]) individuals.  $EC_{50}$  values were estimated using a four-parameter log-logistic function with the plasma dilution factors. Statistically significant FDR values are indicated for comparisons within the fitted linear mixed-effects model (ns indicates non-significant FDR using 0.05 as the cut-off). Tukey's box and whisker plots were used; box limits: interquartile range (IQR); middle line: median; vertical lines: data range (1st quartile  $- 1.5$  IQR; 3rd quartile  $+ 1.5$  IQR). Dots represent each serum value per mutant tested and values from the same serum for the different mutants are connected by gray lines.

lineage enhance cell entry in an in vitro system, consistent with the phenotypic impact of Spike mutations in other lineages of concern<sup>50,51</sup>.

**Spike protein mutations in the Eta lineage reduce antibody neutralization after natural infection.** One of the biggest concerns for any new SARS-CoV-2 variant is the potential impact of accumulated Spike mutations on antibody-mediated neutralization after vaccination or natural infection. To examine the susceptibility of Eta lineage virus Spike proteins to neutralizing antibodies, HIV-1 luciferase reporter viruses were pseudotyped with a subset of the mutations tested above: the original D614 Spike protein, Spike protein containing the D614G mutation, or Spike protein containing the eight mutations that characterize the Eta lineage. Viruses were normalized by p24 concentration and incubated with sera isolated from individuals inoculated with two doses of the BNT162b2 vaccine (Pfizer,  $n = 4$ ), individuals inoculated with two doses of the mRNA-1273 vaccine (Moderna,  $n = 4$ ), or convalescent individuals recovered from COVID-19 infection ( $n = 9$ ) (Supplemental Table 1). Serum from vaccinated individuals was collected roughly two weeks after the second vaccine dose had been administered. Convalescent serum had been collected from residents of Oyo state, Nigeria prior to the emergence of the Eta lineage (from 18 May 2020 to 27 June 2020) at least 5 days, but no later than 5 weeks, after diagnosis with SARS-CoV-2 infection. Antibodies against Spike RBD and NTD were quantified by ELISA in all serum specimens to confirm the presence of an immune response prior to neutralization testing (Supplementary Fig. 8c). On average, serum from vaccinated individuals carried higher antibody titers against both Spike RBD and NTD than convalescent serum. Concentrations of Spike RBD and NTD antibodies in all individuals showed a strong correlation with each other (Spearman  $\rho = 0.93$ ; Supplementary Fig. 8d). After incubation in different dilutions of serum, pseudotyped viruses expressing each Spike mutation were used to challenge HeLa-ACE2 cells with infection measured by luciferase activity in cell lysates after 48 h. The half-maximal effective concentration ( $EC_{50}$ ) was subsequently extrapolated from each dilution curve (Supplementary Fig. 9a) normalizing for differences in Spike infectivity (Supplementary Fig. 9b).

Serum from individuals fully vaccinated with either mRNA vaccine was highly protective against all Spike variants tested, with no significant differences observed between neutralization of viruses expressing the original D614 Spike, G614 Spike, or Eta Spike (Fig. 5c). Convalescent sera similarly demonstrated nearly complete neutralization of viruses carrying D614 and G614 Spike variants when compared to controls without serum treatment. However, the ability of convalescent sera to neutralize Eta Spike (median plasma dilution  $EC_{50} = 2.8 \times 10^{-3}$ ) was significantly reduced compared to

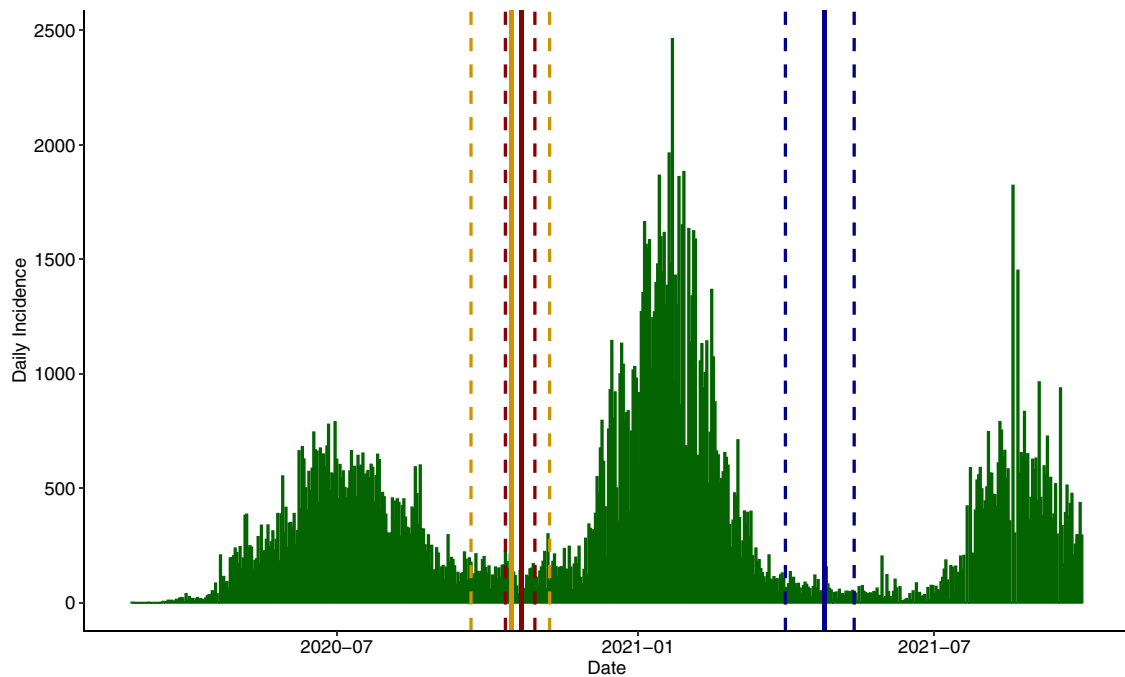
either D614 (median plasma dilution  $EC_{50} = 8 \times 10^{-4}$ ) or G614 Spike (median plasma dilution  $EC_{50} = 4 \times 10^{-4}$ ) (Fig. 5c). Overall Spike RBD and NTD antibody levels in each specimen correlated with the neutralization activity of D614 Spike (Supplementary Fig. 8e), suggesting that increased antibody titer following vaccination may be one correlate of improved protection. However, these results also suggest that the Eta Spike protein may confer partial escape from antibody-mediated neutralization following natural infection with a different lineage.

## Discussion

These results demonstrate the emergence and subsequent dominance of multiple, distinct SARS-CoV-2 lineages in Oyo state, Nigeria. Notably this included the coincident expansion of the B.1.1.7 Alpha 'variant of concern' and the B.1.525 Eta lineage in late 2020 followed by expansion of the uncommon Delta AY.36 lineage of concern by the summer of 2021. As variants with documented evidence of enhanced infectivity, transmissibility, and/or immune evasion continue to emerge and expand across the globe, updated and continued guidance is needed to ensure improved diagnostic and prevention strategies are implemented as soon as possible. The ability of Eta to outcompete Alpha in West Africa, coupled with the evidence of increased infectivity in vitro and enhanced immune evasion after natural infection, suggest that Eta may have warranted designation as a 'variant of concern' had its expansion potential and phenotypic characteristics been recognized earlier. More broadly, this work demonstrates the critical need for international cooperation in infectious disease surveillance in undersampled regions for the monitoring and "early-warning" detection of new pathogenic threats.

The Alpha lineage, which was first reported on 21 December 2020 in the U.K., represented a plurality of cases in Oyo state and throughout Nigeria by early January 2021. While this lineage has not been found to influence the efficacy of neutralizing antibodies produced by vaccination, it is considered a variant of concern as it expanded rapidly in multiple places across the globe, was found to be more infectious in vitro, has evidence of increased transmissibility, and can interfere with diagnostic tests probing over the 69–70 deletion in the S protein<sup>28,52–54</sup>. Compared to the Alpha lineage, many fewer representative cases of the Eta lineage have been reported to date, although it also expanded globally over a similar time frame. This may be due in part to under-sampling of the regions where Eta was most prevalent, including Nigeria and the rest of West Africa. Indeed, phylogenetic analyses of the available sequences are consistent with this region being the source of most Eta cases observed elsewhere globally. That being said, we cannot definitively conclude that the Eta lineage originated in Nigeria due to low sampling and the observation that several sequences from Turkey root outside of the node with the





**Fig. 6 Daily SARS-CoV-2 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 (dark red), B.1.525 (gold), and Delta (dark blue) lineages estimated using Bayesian methods is indicated.

highest statistical support corresponding to TMRCA for the Eta lineage in Nigeria. Importantly, the Eta lineage carries several Spike mutations that have been previously linked to transmissibility and decreased vaccine efficacy, leading to it being designated a ‘variant of interest’ by the WHO and CDC shortly after its description. Not only does the Eta variant share two Spike deletions with Alpha (69-70del and 144del), but this lineage also contains the Spike E484K mutation, which prior studies had suggested could impact immune recognition and vaccine efficacy<sup>55–60</sup>.

The Eta lineage showed evidence of outcompeting the Alpha lineage in Nigeria and across West Africa until the introduction and rapid expansion of the Delta lineage in late spring of 2021. Delta became predominant in the country by June of 2021 with the Eta lineage representing only a small proportion of cases, if any, by the fall. A majority of Delta isolates in our study (84%) belonged to the otherwise rare AY.36 lineage that constitutes less than 0.5% of Delta sequences globally. At the time of the study, only 11 countries (including Nigeria) out of 142 listed in GISAID had a majority of cases attributable to a lineage this rare and, of these 11 countries, only Nigeria and Israel reported more than 100 Delta lineage sequences. Besides the Spike mutations common to all Delta lineages, L452R and P681R, AY.36 also carries an additional Spike mutation, V1104L, in the C-terminus of S2. This additional mutation also appears to have arisen independently in other Delta lineages, including AY.20, AY.22, and AY.31, possibly indicating a convergent evolution process rather than founder effect in Nigeria. Additional studies of this mutation are required to determine its impact on Spike function and immunological profile.

Comparing the estimated TMRCA for the Alpha, Eta, and Delta lineages with the daily incidence reported by the Nigerian Centers for Disease Control (Fig. 6), 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 suggests that these lineages may have benefited from founder effects that boosted overall prevalence of each as

case numbers rose. Less than a month prior to the estimated TMRCA for Alpha and Eta lineages, Nigeria reopened international airports to regular air traffic (5 September 2020). This step was accompanied by strong measures to prevent SARS-CoV-2 introduction to the country, including requiring a negative qRT-PCR test within 96 h 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 qRT-PCR test on day 7 of arrival. Despite these rigorous public health measures, these data are consistent with a model wherein these lineages were seeded from international travel followed by local expansion. Indeed, a similar pattern is observed with the Delta variant, which also entered the country at a time of low incidence (Fig. 6). This would be consistent with the near monophyletic origins observed for each of these lineages in Nigeria and may explain the emergence of two different otherwise rare lineages in the country.

While founder effects may have contributed to the predominance of these lineages in Nigeria following their introduction during a nadir in case counts, cell culture entry analyses conducted using Spike-pseudotyped virus particles suggest that the Spike proteins of Alpha, Delta, and now Eta also enhance viral entry. Previous studies have described the incremental improvement in SARS-CoV-2 cell entry through Spike protein mutations in lineages that subsequently expanded internationally. The D614G mutation in Spike is now nearly fixed across the global population and defines the predominant B.1 lineage from which the above lineages derive. This mutation was found to improve viral entry into airway epithelial cells and confer both increased transmissibility and increased fitness<sup>10–15</sup>. Subsequently, multiple further variations in Spike, including the 69-70 deletion in the Alpha and Eta lineages, have been described that improve viral entry even further over the G614 background<sup>27,43,45,51,53,59,61</sup>. The 69-70 deletion, as well as Q52R to a lesser degree, significantly increased Spike mediated cell entry in our assays despite being located far from the ACE2 binding site (Fig. 5a, b). A deeper examination of this region is needed to understand

possible effects of these mutations on receptor binding or the mechanics of viral entry.

In addition to increased efficiency of viral entry, we also observed decreased antibody neutralization of viruses carrying Eta Spike proteins in convalescent serum from individuals living in Oyo state who had previously had COVID-19 infections. Conversely, vaccinated sera appeared to confer similar protection in concordance to other studies where they have only observed modest neutralization reductions<sup>62</sup>. Other SARS-CoV-2 lineages have also been shown to have increased potential for immune evasion in convalescent sera<sup>60</sup>, including the Beta lineage, which shares a Spike E484K mutation with Eta<sup>30</sup>. The E484K mutation has been shown to arise spontaneously in viruses passaged in convalescent serum<sup>63</sup>, further supporting its role in potentially protecting the virus from antibody responses against other non-variant Spike proteins<sup>64,65</sup>. Moreover, most of the S gene mutations acquired by Eta are located in the NTD of the Spike protein, possibly affecting a known antigenic site<sup>66,67</sup>. This observation suggests likely additional effects of this set of mutations on natural immunity, specifically on antibodies directed at the NTD of the protein<sup>68</sup>. As suggested by our neutralization data, these mutations could confer selective advantage against previously developed natural immunity that could in part explain the selective advantage of Eta in an environment with limited SARS-CoV-2 spread such as Nigeria, where selective advances might play an even more important role.

While the introduction and spread of these lineages during a period of very low incidence, together with their possible fitness advantages, likely facilitated the rise and dominance of these distinct lineages in Nigeria, it is less clear what drove the decline of case counts in this region following each successive wave. The implementation of stricter public health measures, increased vaccination rates, and improved weather conditions have all been considered to influence waning case counts in different countries at different times, but none of these factors are sufficient to explain the resolution of different waves in Nigeria. It is possible that increased public awareness and/or changes in individual behavior can account for some of the decline, but more study of behavioral trends in the region would be needed to investigate these associations.

In sum, the successive expansion and dominance of these two otherwise rare lineages in Nigeria, Eta and AY.36, suggest that distinct viral population dynamics are underlying the epidemic in West Africa. These results underline the critical importance of improving genomic surveillance efforts to better understand and monitor new variants as they arise in distinct regions of the globe. Additional sequencing and surveillance across these under-sampled regions, and elucidation of the biological and clinical significance of new variants, will be required to better understand and prevent emerging variants with increased fitness on a global scale that could threaten vulnerable health systems of West African countries that have avoided thus far the devastating effects of SARS-CoV-2 observed in other parts of the world.

## 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 RNase P probes for sample quality control as previously described (IDT, cat. no. 10006713). All specimens that failed to amplify the RNase P 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. qRT-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. Ct values were calculated using QuantStudio Design and Analysis v1.5.1.

**cDNA synthesis and viral genome amplification.** cDNA synthesis was performed with SuperScript IV First Strand Synthesis Kit (ThermoFisher, cat. no. 18091050) using 11  $\mu$ l 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  $\mu$ l of cDNA in a 25  $\mu$ l 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<sup>69</sup> 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<sup>70</sup> and manually edited using MEGAX v10.1.8<sup>71</sup>. All Maximum Likelihood (ML) phylogenies were inferred with IQ-Tree v2.0.5<sup>72</sup> using its ModelFinder function<sup>73</sup> 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)<sup>74</sup> and with ultrafast bootstrap (UFboot)<sup>75</sup> with 1000 replicates each. TreeTime v0.7.6<sup>76</sup> 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. For B.1.1.7 and Delta global ML phylogenies, due to the high number of sequences available, we performed a temporal subsampling using genome-sampler<sup>77</sup> and sampling 100 sequences per week since the appearance of each lineage. For the B.1.525 global analysis all available 2329 high-quality sequences (high coverage and low N number) from GISAID were used.

Bayesian time-scaled phylogenetic analyses were performed for the phylogenies of B.1.1.7, B.1.525, and Delta lineages separately using a smaller subset (for computational time reasons) of global sequences from GISAID additionally subsampled by diversity in relation to the respective sequences collected from Nigeria using an identity threshold of 99.5% in genome-sampler. Due to the smaller number of B.1.525 sequences available, we included 5 B.1 global sequences subsampled by diversity, as well as the reference sequence (NC\_045512) to help root the phylogenies. We used BEAST v2.5.2<sup>78</sup> to estimate the date and location of the most recent common ancestors (MRCAs) 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 ( $8 \times 10^{-4}$ ) as the prior for the mean, and a standard deviation of 0.004 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  $6.3 \times 10^{-4}$  for the B.1.1.7 inference  $7.2 \times 10^{-4}$  for the B.1.525 analysis and  $6.4 \times 10^{-4}$  for Delta. All the analyses indicated support for a relaxed clock as the average rate coefficient of variation was 0.6 for B.1.1.7, 0.5 for B.1.525, and 0.8 for Delta, 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 5000 steps were computed. The convergence of MCMC chains was monitored using Tracer v1.7.1, ensuring that the effective sample size (ESS) values were >200 for each parameter estimated.

**Cell lines.** To produce HeLa cells (obtained from AIDS reagents) overexpressing Ace2, HEK-293T cells were transfected with pSPAX2, pCMV-VSVG, and pRRL.sin.cPPT.SFFV/Ace2.IRES-puro.WPRE (a gift from Caroline Goujon; Addgene plasmid # 145839; <http://n2t.net/addgene:145839>; RRI-D:Addgene\_145839). HeLa cells were then transduced with the lentiviral particles

and puromycin (InvivoGen) selected as in Mamede et al.<sup>79</sup>. Overexpression was validated by western blotting.

**Generation of SARS-CoV-2 pseudovirus.** HEK-293T cells were transfected with a 3:2 ratio of NL4-3-nanoluc delta env plasmid (gift from Dr. Thomas Hope) and pCAGGS-Spike Sars-CoV2 plasmid (and introduced mutations) respectively. (The following reagent was produced under HHSN272201400008C and obtained through BEI Resources, NIAID, NIH: Vector pCAGGS Containing the SARS-Related Coronavirus 2, Wuhan-Hu-1 Spike Glycoprotein Receptor Binding Domain (RBD), NR-52309). Culture media was changed 16 h post transfection and viral particles were harvested at 48 h. Viral particles were concentrated using a 20% sucrose gradient with overnight centrifugation at 5600 rcf at 4 °C and then resuspended into fresh media at a 500x concentration. The concentrated Spike pseudoviruses were quantified using HIV-1 Gag p24 Quantikine ELISA Kit (R&D Systems), and the infectivity was determined using Hela-Ace2 cells and the Nano-Glo® Dual-Luciferase® Reporter Assay (Promega). Spike protein production was verified by western blot, using concentrated viral particles combined with laemmli buffer containing 2-mercaptoethanol, heated to 98° for 20 min. Sample concentration was standardized by p24 ELISA and samples were run on a 4–20% tris-glycine gel prior to being transferred to a PVDF membrane. The membrane was probed with Anti-SARS-CoV-2 Spike Protein S1/S2 (Invitrogen PA5-112048) and anti-p24 mAb (NIH-ARP 3537).

**Synthesis of B.1.525 lineage mutations.** The introduction of the B.1.525 lineage mutations into pCAGGS-Spike Sars-CoV2 plasmids (BEI) was performed by restriction digest of the plasmid, followed by NEBuilder HiFi DNA assembly (New England Biolabs) with either amplified by PCR products or synthesized DNA fragments (GeneWiz). The full modified sequences were verified and confirmed by Sanger sequencing (Supplementary Methods).

**Antibody neutralization assay.** HeLa-Ace2 cells were plated in 96-well black wall plates at 10<sup>5</sup> cells per well and incubated at 37 °C, with 5% CO<sub>2</sub> for 24 hours. Media was then removed, and serum was added using serial dilutions. Pseudotyped virus (pCAGGS-Spike Sars-CoV2 and mutants; NL4-3-nanoluc delta env) was added to each well, and cells were returned to the incubator. After a 48-hour incubation, reporter gene expression was determined with Nano-Glo® Dual-Luciferase® Reporter Assay (Promega) following the vendor's protocol using a luminometer (Cytation3, Biotek). We estimated the amount of RBD-binding and NTD-binding antibodies at half-maximal effective concentration (EC<sub>50</sub>) for each serum sample by fitting a four-parameter log-logistic function to each dilution curve for both RBD and NTD using the drc v3.0-1 package in R v4.0.3. For this estimation, we used the luminescence values measured in relative light units (RLU) normalized both by the values obtained with the negative controls without virus and sera and the positive controls with each corresponding virus and no sera added.

**RBD- and NTD-binding by ELISA.** Ni-NTA HisSorb 96-well plates (Qiagen) were coated with 50 µl of either His-tagged RBD (Acro Biosystems, SPD-S52H6, AA 306-527) or His-tagged NTD (Acro Biosystems, S1D-C52H6, AA13-303) at a concentration of 2 µg/ml overnight at 4 °C; all subsequent steps were performed at room temperature. Plates were blocked with 3% milk; plasma was diluted 1:500 in PBS for RBD ELISA and 1:1500 for NTD ELISA prior to adding 100 µl per well for 2 h. Plates were washed with PBS-T three times, followed by incubation with secondary anti-human IgG Fc HRP (Southern Biotech) at 1:4000 for 1 h. Plates were washed three additional times with PBS-T, followed by addition of TMB substrate for 10 min. The reaction was stopped with the addition of 3 M HCl; and plates were read at an OD 450. Anti-RBD antibody CR3022 (Creative Biolabs) was used to make a standard curve with 1:2 dilutions in the range of 0.5 µg/ml to 0.03 µg/ml, utilizing a 4-parameter fit to interpolate sample concentrations. Anti-NTD antibody SPD-M121 (Acro Biosystems) was used to make a standard curve with 1:2 dilutions in the range of 0.5 µg/ml to 0.03 µg/ml, utilizing a 4-parameter fit to interpolate sample concentrations.

**Structural modeling of B.1.525 protein.** A model of the B.1.525 NTD domain, with the Q52R and A67V mutations and H69, V70 and Y144 deletions was built using SwissModel<sup>80</sup> using a PDB coordinate of the Fab C25 complex with the NTD of SARS-CoV-2 (PDB 7m8j)<sup>81</sup> as a template. The complex was minimized and binding energy was estimated using the EvoEF program, part of the EvoDesign pipeline<sup>82,83</sup>.

**Statistics.** All statistical analyses were performed in R v 4.0.3. Statistical analysis of cell entry and neutralization EC<sub>50</sub> were performed fitting Linear Mixed-Effects Models using lme4 v1.1-27 package. For cell entry, we tested for significant differences in cell entry between mutants with increasing concentrations of pseudotyped viruses including the interaction between mutant and p24 concentration in the model and accounting for experiment and plate effects. For neutralization EC<sub>50</sub> differences we included in the model the different groups of sera, each mutant, as well as the interaction of both parameters, with the addition of the possible effect of within-donor sera correlation. Afterwards we performed every possible contrast

within each model and used FDR to assess significance while controlling for multiple comparisons. A linear model including patient sex and age as confounders was fitted to test for differences in Ct value between lineages.

## Data availability

The 378 complete SARS-CoV-2 genomes from Nigeria and the 37 from Mali were deposited in the public databases GISAID and GenBank (Accession numbers for both databases are provided in Supplementary Data 1). The publicly available data sources used for this study are as follows: Global sequences were obtained from GISAID database (<https://www.gisaid.org/>; Supplementary Data 2); nextstrain.org was used for clade information and analysis; cov-lineages.org was used for lineage information and analysis; epidemiological data from Nigeria was obtained from the Nigerian Centers for Disease Control (<https://covid19.ncdc.gov.ng/>) and Johns Hopkins Coronavirus resource center (<https://coronavirus.jhu.edu/>); additional information about lineage distribution and dynamics was obtained from outbreak.info, covariants.org, and coronavirus3d.org. Nigeria's map was created using R packages rnatuarearth v0.1.0 and sp v1.4.6. Source data are provided with this paper.

Received: 10 November 2021; Accepted: 18 January 2022;

Published online: 03 February 2022

## References

- Wang, C., Horby, P. W., Hayden, F. G. & Gao, G. F. A novel coronavirus outbreak of global health concern. *Lancet* **395**, 470–473 (2020).
- Zhu, N. et al. A novel coronavirus from patients with pneumonia in China, 2019. *New Engl. J. Med.* **382**, 727–733 (2020).
- Zhou, P. et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* **579**, 270–273 (2020).
- van Dorp, L. et al. Emergence of genomic diversity and recurrent mutations in SARS-CoV-2. *Infect. Genet. Evol.* **83**, 104351 (2020).
- Li, X. et al. Transmission dynamics and evolutionary history of 2019-nCoV. *J. Med. Virol.* **92**, 501–511 (2020).
- Meredith, L. W. 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.* **20**, 1263–1271 (2020).
- Dearlove, B. et al. A SARS-CoV-2 vaccine candidate would likely match all currently circulating variants. *Proc. Natl Acad. Sci. USA* **117**, 23652–23662 (2020).
- Lu, J. et al. Genomic epidemiology of SARS-CoV-2 in Guangdong province, China. *Cell* **181**, 997–1003.e1009 (2020).
- Gonzalez-Reiche, A. S. et al. Introductions and early spread of SARS-CoV-2 in the New York City area. *Science* **369**, 297–301 (2020).
- Lorenzo-Redondo, R. et al. A clade of SARS-CoV-2 viruses associated with lower viral loads in patient upper airways. *EBioMedicine* **62**, 103112 (2020).
- 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.e819 (2020).
- Hou, Y. J. et al. SARS-CoV-2 D614G variant exhibits efficient replication ex vivo and transmission in vivo. *Science* **370**, 1464–1468 (2020).
- Li, Q. et al. The impact of mutations in SARS-CoV-2 spike on viral infectivity and antigenicity. *Cell* **182**, 1284–1294.e1289 (2020).
- Volz, E. et al. Evaluating the effects of SARS-CoV-2 spike mutation D614G on transmissibility and pathogenicity. *Cell* **184**, 64–75.e11 (2021).
- Plante, J. A. et al. Spike mutation D614G alters SARS-CoV-2 fitness. *Nature* **592**, 116–121 (2021).
- Centers for Disease Control and Prevention. *SARS-CoV-2 Variant Classifications and Definitions* (2021).
- Konings, F. et al. SARS-CoV-2 variants of interest and concern naming scheme conducive for global discourse. *Nat. Microbiol.* **6**, 821–823 (2021).
- Lauring, A. S. & Hodcroft, E. B. Genetic variants of SARS-CoV-2-what do they mean? *JAMA* **325**, 529–531 (2021).
- Grubaugh, N. D., Hodcroft, E. B., Fauver, J. R., Phelan, A. L. & Cevik, M. Public health actions to control new SARS-CoV-2 variants. *Cell* **184**, 1127–1132 (2021).
- Davies, N. G. et al. Estimated transmissibility and severity of novel SARS-CoV-2 Variant of Concern 202012/01 in England. *medRxiv* <https://doi.org/10.1101/2020.12.24.20248822> (2020).
- U.S. Food and Drug Administration Administration. FDA authorizes revisions to fact sheets to address SARS-CoV-2 variants for monoclonal antibody products under emergency use authorization (2021).
- Edara, V. V. et al. Reduced binding and neutralization of infection- and vaccine-induced antibodies to the B.1.351 (South African) SARS-CoV-2 variant. *bioRxiv* <https://doi.org/10.1101/2021.02.20.432046> (2021).
- Xie, X. et al. Neutralization of SARS-CoV-2 spike 69/70 deletion, E484K and N501Y variants by BNT162b2 vaccine-elicited sera. *Nat. Med.* **27**, 620–621 (2021).

24. Wang, L. et al. Antibodies with potent and broad neutralizing activity against antigenically diverse and highly transmissible SARS-CoV-2 variants. *bioRxiv* <https://doi.org/10.1101/2021.02.25.432969> (2021).
25. Tegally, H. et al. Sixteen novel lineages of SARS-CoV-2 in South Africa. *Nat. Med.* **27**, 440–446 (2021).
26. Leung, K., Shum, M. H., Leung, G. M., Lam, T. T. & Wu, J. T. Early transmissibility assessment of the N501Y mutant strains of SARS-CoV-2 in the United Kingdom, October to November 2020. *Euro Surveill.* **26**, 2002106 (2021).
27. Volz, E. et al. Transmission of SARS-CoV-2 Lineage B.1.1.7 in England: Insights from linking epidemiological and genetic data. *medRxiv* <https://doi.org/10.1101/2020.12.30.20249034> (2021).
28. Rambaut, A. et al. Preliminary genomic characterisation of an emergent SARS-CoV-2 lineage in the UK defined by a novel set of spike mutations. *nCoV-2019 Genomic Epidemiology* (18 Dec 2020).
29. Rambaut, A. et al. A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. *Nat. Microbiol.* **5**, 1403–1407 (2020).
30. Faria, N. R. et al. Genomics and epidemiology of the P.1 SARS-CoV-2 lineage in Manaus, Brazil. *Science* **372**, 815–821 (2021).
31. Centers for Disease Control and Prevention. *SARS-CoV-2 Variant Classifications and Definitions* (2021).
32. Lemey, P. et al. Accommodating individual travel history and unsampled diversity in Bayesian phylogeographic inference of SARS-CoV-2. *Nat. Commun.* **11**, 5110 (2020).
33. Wilkinson, E. et al. A year of genomic surveillance reveals how the SARS-CoV-2 pandemic unfolded in Africa. *Science* **374**, 423–431 (2021).
34. 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* **6**, e004408 (2021).
35. Inzaule, S. C., Tessema, S. K., Kebede, Y., Ogwel Ouma, A. E. & Nkengasong, J. N. Genomic-informed pathogen surveillance in Africa: opportunities and challenges. *Lancet Infect. Dis.* **21**, e281–e289 (2021).
36. Centers for Disease Control and Prevention. *CDC 2019–Novel Coronavirus (2019-nCoV) Real Time RT-PCR Diagnostic Panel* (2020).
37. Quick, J. et al. Multiplex PCR method for MinION and Illumina sequencing of Zika and other virus genomes directly from clinical samples. *Nat. Protoc.* **12**, 1261–1276 (2017).
38. Quick, J. nCoV-2019 sequencing protocol v2 (GunIt).
39. Elbe, S. & Buckland-Merrett, G. Data, disease and diplomacy: GISAID's innovative contribution to global health. *Glob. Chall.* **1**, 33–46 (2017).
40. Shu, Y. & McCauley, J. GISAID: Global initiative on sharing all influenza data - from vision to reality. *Euro Surveill.* **22**, 30494 (2017).
41. Nigeria Centre for Disease Control. COVID-19 NIGERIA. NCDC Coronavirus COVID-19 Microsite (2022).
42. O'Toole, A. & Hill, V. Global Report of B.1.525 lineage (2021).
43. Frampton, D. et al. Genomic characteristics and clinical effect of the emergent SARS-CoV-2 B.1.1.7 lineage in London, UK: a whole-genome sequencing and hospital-based cohort study. *Lancet. Infect. Dis.* **21**, 1246–1256 (2021).
44. Meng, B. et al. Recurrent emergence of SARS-CoV-2 spike deletion H69/V70 and its role in the variant of concern lineage B.1.1.7. *Cell Rep.* **35**, 109292 (2021).
45. Collier, D. A. et al. Sensitivity of SARS-CoV-2 B.1.1.7 to mRNA vaccine-elicited antibodies. *Nature* **593**, 136–141 (2021).
46. Cerutti, G. et al. Potent SARS-CoV-2 neutralizing antibodies directed against spike N-terminal domain target a single supersite. *Cell Host Microbe* **29**, 819–833 e817 (2021).
47. Liu, L. et al. Potent neutralizing antibodies against multiple epitopes on SARS-CoV-2 spike. *Nature* **584**, 450–456 (2020).
48. Rogers, T. F. et al. Isolation of potent SARS-CoV-2 neutralizing antibodies and protection from disease in a small animal model. *Science* **369**, 956–963 (2020).
49. Xia, S. et al. The role of furin cleavage site in SARS-CoV-2 spike protein-mediated membrane fusion in the presence or absence of trypsin. *Signal Transduct. Target Ther.* **5**, 92 (2020).
50. Harvey, W. T. et al. SARS-CoV-2 variants, spike mutations and immune escape. *Nat. Rev. Microbiol.* **19**, 409–424 (2021).
51. McCarthy, K. R. et al. Recurrent deletions in the SARS-CoV-2 spike glycoprotein drive antibody escape. *Science* **371**, 1139–1142 (2021).
52. Xie, X. et al. Neutralization of N501Y mutant SARS-CoV-2 by BNT162b2 vaccine-elicited sera. *bioRxiv* <https://doi.org/10.1101/2021.01.07.425740> (2021).
53. Galloway, S. E. et al. Emergence of SARS-CoV-2 B.1.1.7 Lineage — United States, December 29, 2020–January 12, 2021. *Morbidity Mortal. Wkly Rep.* **70**, 95–99 (2021).
54. Public Health England. Investigation of novel SARS-CoV-2 variant: variant of concern 202012/01, Technical Briefing 3. (2021).
55. Wu, K. et al. mRNA-1273 vaccine induces neutralizing antibodies against spike mutants from global SARS-CoV-2 variants. *bioRxiv* <https://doi.org/10.1101/2021.01.25.427948> (2021).
56. Wang, Z. et al. mRNA vaccine-elicited antibodies to SARS-CoV-2 and circulating variants. *bioRxiv* <https://doi.org/10.1101/2021.01.15.426911> (2021).
57. Wibmer, C. K. et al. SARS-CoV-2 501Y.V2 escapes neutralization by South African COVID-19 donor plasma. *bioRxiv* <https://doi.org/10.1101/2021.01.18.427166> (2021).
58. Weisblum, Y. et al. Escape from neutralizing antibodies by SARS-CoV-2 spike protein variants. *Elife* **9**, e61312 (2020).
59. Zhou, D. et al. Evidence of escape of SARS-CoV-2 variant B.1.351 from natural and vaccine induced sera. *Cell* **184**, 2348–2361.e6 (2021).
60. Greaney, A. J. et al. Comprehensive mapping of mutations in the SARS-CoV-2 receptor-binding domain that affect recognition by polyclonal human plasma antibodies. *Cell Host Microbe* **29**, 463–476.e466 (2021).
61. Bo Meng, S. A. K. et al. Recurrent emergence of SARS-CoV-2 spike deletion H69/V70 and its role in the variant of concern lineage B.1.1.7. *Cell Rep.* **35**, 109292 (2021).
62. Liu, J. et al. BNT162b2-elicited neutralization of B.1.617 and other SARS-CoV-2 variants. *Nature* **596**, 273–275 (2021).
63. Andreano, E. et al. SARS-CoV-2 escape in vitro from a highly neutralizing COVID-19 convalescent plasma. *bioRxiv* <https://doi.org/10.1101/2020.12.28.424451> (2020).
64. Baum, A. et al. Antibody cocktail to SARS-CoV-2 spike protein prevents rapid mutational escape seen with individual antibodies. *Science* **369**, 1014–1018 (2020).
65. Liu, Z. et al. Identification of SARS-CoV-2 spike mutations that attenuate monoclonal and serum antibody neutralization. *Cell Host Microbe* **29**, 477–488 e474 (2021).
66. Suryadevara, N. et al. Neutralizing and protective human monoclonal antibodies recognizing the N-terminal domain of the SARS-CoV-2 spike protein. *Cell* **184**, 2316–2331.e2315 (2021).
67. McCallum, M. et al. N-terminal domain antigenic mapping reveals a site of vulnerability for SARS-CoV-2. *Cell* **184**, 2332–2347.e2316 (2021).
68. Lok, S. M. An NTD supersite of attack. *Cell Host Microbe* **29**, 744–746 (2021).
69. Grubaugh, N. D. et al. An amplicon-based sequencing framework for accurately measuring intrahost virus diversity using PrimalSeq and iVar. *Genome Biol.* **20**, 8 (2019).
70. Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol. Biol. Evol.* **30**, 772–780 (2013).
71. Kumar, S., Stecher, G., Li, M., Nknyaz, C. & Tamura, K. MEGA X: molecular evolutionary genetics analysis across computing platforms. *Mol. Biol. Evol.* **35**, 1547–1549 (2018).
72. 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).
73. Kalyaanamoorthy, S., Minh, B. Q., Wong, T. K. F., von Haeseler, A. & Jermini, L. S. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat. Methods* **14**, 587–589 (2017).
74. Anisimova, M. & Gascuel, O. Approximate likelihood-ratio test for branches: a fast, accurate, and powerful alternative. *Syst. Biol.* **55**, 539–552 (2006).
75. Hoang, D. T., Chernomor, O., von Haeseler, A., Minh, B. Q. & Vinh, L. S. UFBoot2: improving the ultrafast bootstrap approximation. *Mol. Biol. Evol.* **35**, 518–522 (2018).
76. Sagulenko, P., Puller, V. & Neher, R. A. TreeTime: maximum-likelihood phylodynamic analysis. *Virus Evol.* **4**, vex042 (2018).
77. Bolyen, E. et al. Reproducibly sampling SARS-CoV-2 genomes across time, geography, and viral diversity. *PLoS Comput. Biol.* **16**, e1006650 (2020).
78. Bouckaert, R. et al. BEAST 2.5: an advanced software platform for Bayesian evolutionary analysis. *PLoS Comput. Biol.* **15**, e1006650 (2019).
79. Mamede, J. I., Sitbon, M., Battini, J. L. & Courgnaud, V. Heterogeneous susceptibility of circulating SIV isolate capsids to HIV-interacting factors. *Retrovirology* **10**, 77 (2013).
80. Waterhouse, A. et al. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res.* **46**, W296–W303 (2018).
81. Voss, W. N. et al. Prevalent, protective, and convergent IgG recognition of SARS-CoV-2 non-RBD spike epitopes. *Science* **372**, 1108–1112 (2021).
82. Huang, X., Pearce, R. & Zhang, Y. EvoEF2: accurate and fast energy function for computational protein design. *Bioinformatics* **36**, 1135–1142 (2020).
83. Pearce, R., Huang, X., Setiawan, D. & Zhang, Y. EvoDesign: designing protein-protein binding interactions using evolutionary interface profiles in conjunction with an optimized physical energy function. *J. Mol. Biol.* **431**, 2467–2476 (2019).
84. Alaa Abdel Latif, J. L. M. et al. *The Center for Viral Systems Biology. Lineage Comparison* (outbreak.info, 2021).
85. Benton, D. J. et al. Receptor binding and priming of the spike protein of SARS-CoV-2 for membrane fusion. *Nature* **588**, 327–330 (2020).

## 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. Genome sequencing and analysis were supported by resources and staff at the Center for Pathogen Genomics and Microbial Evolution in the Northwestern University Havy Institute for Global Health. Funding for this work was provided by: a Northwestern Havy 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 University Center for Advanced Technologies (NUCATS - J.F.H.); a COVID-19 Collaborative Innovation Award from the NUCATS (R.L.R.); a CTSA supplement to NCATS UL1 TR002389 (J.F.H., E.A.O., and 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 R21 AI163912 (J.F.H.); NIH grant U19 AI135964 (E.A.O.); NIH grant D43 TW009608 (B.O.T. and O.M.A.); NIH Fogarty International Center award D43TW009608 (B.O.T.); NIH grant R35 GM118187 (A.G.) and NIH-NIAID Center for Structural Genomics of Infectious Diseases award HHSN272201700060C (A.G.); USAID Bureau for Resilience and Food Security/Center for Agriculture-led Growth, Cooperative Agreement 720-OAA-18-LA-00003 as part of Feed the Future Innovation Lab for Collaborative Research on Sustainable Intensification (R.L.R.); and through a generous contribution from the Walder Foundation Foundation's Chicago Coronavirus Assessment Network (Chicago CAN) Initiative (J.F.H., E.A.O., and R.L.R.) and Walder Foundation grant numbers SCI16 (J.R.S.) and 21-00147(J.R.S.). The authors also acknowledge the Nigeria Center for Disease Control (NCDC) for providing leadership and infrastructure that facilitated the collection of samples included in this study.

## Author contributions

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

## Competing interests

The authors declare no competing interests.

## Ethics Approval

This study was reviewed and approved by the National Health Research Ethics Committee of Nigeria (NHREC) with NHREC protocol number NHREC/01/01/2007 and the University of Ibadan/University College Hospital Health Research Ethics Committee (UI/UCH HREC) with protocol number UI/EC/20/0233. Project name: Phylodynamics and molecular epidemiology of SARS-CoV-2 in Nigeria. Samples from convalescent individuals from Nigeria were collected under a study approved by NHREC with NHREC approval number IRB/20/028. Project name: Plasma and PBMC repository to

elucidate the immune and cytokine profile of SARS-CoV-2 Infected Nigerians. Samples from vaccinated individuals from the USA were collected under a study approved by the Institutional Review Board of Rush University with protocol number 20032502-IRB01. Project name: SARS CoV-2 Pathogenesis, Immunity, and drug Development. The Mali sample collection was approved by the Ethics Committee of the Faculty of Medicine and Odontostomatology and the Faculty of Pharmacy (FMOS/FAPH) of the University of Sciences, Techniques and Technologies of Bamako (USTTB) in Mali under the number 2020/193/CE/FMOS/FAPH. Project Name: Clinical and Virological Characterization of the COVID-19 pandemic in Mali. Residual nasopharyngeal swab samples from diagnostic tests administered in Nigeria were provided to the investigators in a de-identified manner with no additional clinical data and were used solely for the purposes of SARS-CoV-2 genomic surveillance. The ethics approval for the use of residual nasopharyngeal swab samples included a waiver of informed consent.

## Informed consent

Informed consent was obtained for all the sera samples used in this study for neutralization experiments, as well as all samples collected in Mali for viral sequencing. Only sera samples from adult participants were used for neutralization experiments.

## Additional information

**Supplementary information** The online version contains supplementary material available at <https://doi.org/10.1038/s41467-022-28317-5>.

**Correspondence** and requests for materials should be addressed to Judd F. Hultquist or Ramon Lorenzo-Redondo.

**Peer review information** *Nature Communications* thanks the anonymous reviewer(s) for their contribution to the peer review of this work.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s) 2022