 doi:10.20944/preprints202210.0404.v1

Article

Bayesian Molecular Dating Analyses Combined with Mutational Profiling Suggest an Independent Origin and Evolution of SARS-CoV-2 Omicron BA.1 and BA.2 Sub-lineages

Naveen Kumar[1#*], Rahul Kaushik[2,3#], Ashutosh Singh[1], Vladimir N. Uversky[3,4], Kam Y. J. Zhang[3], Upasana Sahu[1], Sandeep Bhatia[1] and Aniket Sanyal[1]

[1] Diagnostics & Vaccines Group, ICAR- National Institute of High Security Animal Diseases, Bhopal 462022, India; naveen.kumar4@icar.gov.in (N.K.), sashutosh790@gmail.com (A.S.), upasanasahu2pch@gmail.com (U.S.), sbhatia1967@gmail.com (S.B.), aniket.sanyal@gmail.com (A.S.)
[2] Biotechnology Research Center, Technology Innovation Institute, P. O. Box 3692, Masdar City, Abu Dhabi, UAE.; rahul.kaushik@tii.ae (R.K.)
[3] Laboratory for Structural Bioinformatics, Center for Biosystems Dynamics Research, RIKEN, 1-7-22 Suehiro, Yokohama, Kanagawa 230-0045, Japan; rahul.kaushik@riken.jp (R.K.), kamzhang@riken.jp (K.Y.J.Z.)
[3] Department of Molecular Medicine, Morsani College of Medicine, University of South Florida, Tampa, FL 33612, USA; vuversky@usf.edu (V.N.U)
[4] Institute for Biological Instrumentation of the Russian Academy of Sciences, Federal Research Center 'Pushchino Scientific Center for Biological Research of the Russian Academy of Sciences', Moscow region, 142290 Pushchino, Russia; vuversky@usf.edu (V.N.U)
* Correspondence: naveen.kumar4@icar.gov.in (N.K.); Tel.: (+91-7552759204)
# These authors contributed equally to this manuscript.

Abstract: The ongoing evolution of severe acute respiratory syndrome-coronavirus-2 (SARS-CoV-2) has resulted in the recent emergence of a highly divergent variant of concern (VOC) defined as Omicron or B.1.1.529. This VOC is of particular concern because it has the potential to evade most therapeutic antibodies and has undergone a sustained genetic evolution, resulting in the emergence of five distinct sub-lineages. However, the evolutionary dynamics of initially identified Omicron BA.1 and BA.2 sub-lineages remain poorly understood. Herein, we combined Bayesian phylogenetic analysis, mutational profiling, and selection pressure analysis to track virus genetic changes that drive the early evolutionary dynamics of the Omicron. Based on the Omicron dataset chosen for the improved temporal signals and sampled globally between November 2021 and January 2022, most recent common ancestor (tMRCA) and substitution rates for BA.1 were estimated to be 18 September 2021 (95% highest posterior density (HPD) 04 August – 22 October 2021) and 1.435×10^{-3} (95% HPD = 1.021×10^{-3} – 1.869×10^{-3}) substitution/site/year, respectively, whereas 03 November 2021 (95% highest posterior density (HPD) 26 September – 28 November 2021) and 1.074×10^{-3} (95% HPD = 6.444×10^{-4} – 1.586×10^{-3}) substitution/site/year for BA.2 sub-lineage. The findings of this study suggest that the Omicron BA.1 and BA.2 sub-lineages originated independently and evolved over time. Furthermore, we identified multiple sites in spike protein undergoing continued diversifying selection that may alter the neutralization profile of BA.1. This study shed light on the ongoing global genomic surveillance and Bayesian molecular dating analyses to better understand the evolutionary dynamics the virus and, as a result, mitigate the impact of emerging variants on public health.

Keywords: COVID-19; SARS-CoV-2 Omicron; tMRCA; evolutionary rate; mutational profiling; selection pressure

 doi:10.20944/preprints202210.0404.v1

1. Introduction

Since the inception of the coronavirus disease 2019 (COVID-19) pandemic caused by severe acute respiratory syndrome-coronavirus-2 (SARS-CoV-2) in December 2019, multiple variants have emerged and rapidly spread, resulting in significant changes in SARS-CoV-2's global evolutionary dynamics. The current nomenclature schemes for these variants are the Global Initiative on Sharing All Influenza Data (GISAID) [1], Pango [2], and Nextstrain [3]. Nonetheless, the World Health Organization (WHO) has classified these variants into Variants of Interest (VOIs) and Variants of Concern (VOCs), which makes it easier for their tracking, research, and to eventually guide the ongoing COVID-19 pandemic response on a global scale [4]. The former carries genetic changes that may affect COVID-19 severity, transmissibility, and ability to escape from existing diagnostic and therapeutic approaches, resulting in an emerging risk to global public health with increased community transmission. The latter has been shown to carry one or more changes that directly affect global public health, such as enhanced virulence and/or transmissibility, resulting in a significant change in COVID-19 epidemiology and clinical disease severity, reduced effectiveness of accessible diagnostics, therapeutics, vaccines, and social measures.

Since late 2020, a number of VOCs have emerged that are seriously affecting global public health, including Alpha (PANGO lineage, B.1.1.7) in the United Kingdom, Beta (B.1.351) in South Africa, Gamma (P.1) in Brazil, and Delta (B.1.617.2) in India [2,5-7]. While Alpha, Beta, and Gamma VOCs disseminated globally at this time, it was the extremely transmissible Delta variant that ultimately displaced all of the other VOCs in the majority of the countries [8]. In November 2021, while Delta was still causing significant levels of transmission in several countries, a new variant, B.1.1.529 was detected in a specimen collected on 9 November 2021 from South Africa and reported to WHO on 24 November 2021 [9]. The WHO named this variant as the Omicron variant—the fifth VOC of SARS-CoV-2.

Omicron is distantly related to other VOCs, and it has dominated SARS-CoV-2 infections worldwide; however, it has been found to produce less severe disease than previously documented VOCs [10,11]. Nonetheless, epidemiological evidence suggests that Omicron VOC have a 2- to 3-fold greater risk of re-infection, which could be due to the capacity to circumvent the pre-existing immunity acquired by infections or vaccinations [12]. In comparison with the original Wuhan-Hu-1 reference strain, Omicron VOC accumulated 53 mutations, with 30 mutations in the spike protein alone. This unusual preponderance of mutations, particularly in Omicron's spike protein, has been demonstrated to enhance the transmissibility [13,14] as well as to escape from several neutralizing antibodies (NAbs) [15-18].

Omicron VOC is undergoing substantial genetic evolution, as evidenced by the identification of three sub-lineages of Omicron (BA.1, BA.2, and BA.3) concurrently, and these three sub-lineages are quite distinct from one another [9]. Initially, BA.1 proved to be the most prolific sub-lineage and spread rapidly worldwide; however, BA.2 later emerged as the dominant sub-lineage and surpassed BA.1 globally. Recently, two more sub-lineages, BA.4 and BA.5, were discovered in South Africa [19]. Here, we focus especially on the BA.1 and BA.2 sub-lineages, which quickly spread concurrently in several countries throughout the world, and were the first recognized as the Omicron VOC.

Given the scarcity of studies decoding the origin and evolution of Omicron VOC, we first deciphered and compared the unique mutational profiles of Omicron VOCs (BA.1 and BA. 2 sub-lineages) to those of previously known VOCs (Alpha, Beta, Gamma, and Delta). Their functional characterization is also discussed in detail, which will improve our understanding of their clinical implications. Second, we aimed to explore the evolutionary dynamics of SARS-CoV-2 Omicron VOC by utilizing root-to-tip regression analysis, recombination analysis, and Bayesian evolutionary analysis based on 32,170 whole genomes sampled globally between November 2021 and January 2022. Finally, we

 doi:10.20944/preprints202210.0404.v1

characterize the selective pressures that may have acted during the initial rapid evolution of the BA.1 and BA.2 sub-lineages.

2. Materials and Methods

2.1. Collation of SARS-CoV-2 Omicron variant complete genome dataset

The high-quality complete genome sequences of SARS-CoV-2 Omicron variant (n = 32,170) were obtained from the GISAID, which was deposited between November 2021 and January 2022 (http://gisaid.org/) [20]. These sequences were filtered out based on the following criteria: (1) low-quality sequences carrying unusual characters other than A, T, G, and C; (2) duplicate sequences with 100% nucleotide identity; and (3) sequences with incomplete associated information, such as sampling dates. Following filtration, the remaining high-quality sequences (BA.1 = 1770 and BA.2 = 1037) were aligned using MAFFT v.7.490 [21], and subsequently open reading frames (ORFs) were extracted manually from the aligned complete genomic sequences of SARS-CoV-2 Omicron variant using BioEdit [22].

2.2. Recombination analysis

The complete coding genomic sequences of SARS-CoV-2, where the ORFs were concatenated in the following order: ORF1ab + S + ORF3a + E + M + ORF6 + ORF7a + ORF7b + ORF8 + N + ORF10, were screened for recombination signals using RDP v4.101, which implements nine distinct algorithms: RDP, GENECONV, Bootscan, MaxChi, Chimaera, Siscan, PhylPro, LARD, and 3seq [23]. Using the default settings, these sequences were examined for each identified recombination breakpoint. In order to reduce false positive recombination signals, we only considered recombination events detected by at least two of the nine algorithms.

2.3. Root-to-tip regression analysis to assess the temporal signals

In order to assess the temporal signals in the dataset, we employed root-to-tip regression analysis on the entire coding genomic sequences of SARS-CoV-2 Omicron variant. In summary, root-to-tip regression analyses are commonly used to estimate the relationship between root-to-tip genetic divergence and sampling dates generated from Maximum Likelihood (ML) phylogeny. The slope of the regression line gives an estimate of the evolutionary rates (substitutions per site per year), whereas the intercept with the time axis estimates the age of the root.

We first screened for the Maximum Likelihood (ML) fits of 88 alternative nucleotide substitution models for the SARS-CoV-2 Omicron sub-lineages (BA.1 and BA.2) dataset, and subsequently used ModelFinder to identify the best fitting nucleotide substitution model based on the Bayesian Information Criterion (BIC) (Table S1) [24]. The phylogenetic trees were then estimated using ML inference with ultrafast bootstrap with 1000 replicates as implemented in IQ tree v2.1.2 [25]. Finally, these ML trees were used to investigate the temporal molecular evolutionary signals for each SARS-CoV-2 Omicron sub-lineage using TempEst v1.5.3 [26].

2.4. Molecular clock phylogenetics

To infer the substitution/evolutionary rates and timescale of SARS-CoV-2 Omicron variant, Bayesian inference analyses was performed on datasets containing 161 dated non-recombinant nucleotide sequences for the complete coding sequences of BA.1 (n = 75) and BA.2 (n = 86) sub-lineages using a Markov Chain Monte Carlo (MCMC) framework [27], implemented in Bayesian evolutionary analysis by sampling trees (BEAST) v2.6.7 [28]. As previously indicated, the best-fit nucleotide substitution model for each Omicron sub-lineage dataset was chosen.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 October 2022 doi:10.20944/preprints202210.0404.v1

To identify the best combination of tree priors and clock models, we tested and compared four coalescent tree priors: a constant population size [29], exponential population [30], Bayesian skyline [31], and extended Bayesian skyline [32] tree priors, and two clock models: strict clock and an uncorrelated relaxed clock with log-normal distribution (UCLN) [33]. All Bayesian analyses were run for 100 million steps across two independent MCMC simulations with states and parameters sampled after every 10,000 steps. To find the best tree prior-clock model combination, Bayesian model testing, which as a statistical fit measure calculated by computing the log marginal likelihood, was performed and subsequently ranked each model combination accordingly. The log Bayes factor (BF) is the difference between two tree prior-clock models' log marginal likelihoods [34]. A log BF of at least 1.1 in favor of a model is described as 'substantial evidence', with 2.3 being 'strong' and 4.6 being 'decisive' [35]. We considered two marginal likelihood estimators: path sampling and stepping-stone sampling [36-38]. The parameters for the best-fit model combination for Omicron sub-lineages attained an effective sample size of more than 200, indicating adequate sampling. Using Tracer v1.7.1, we extracted time to the most recent common ancestor (tMRCA) and clock rate estimations from the best-fit model combination [39]. After deleting the first 10% of samples as burn-in, the maximum clade credibility (MCC) tree was extracted using TreeAnnotator v1.8.4 [40]. The MCC trees were visualized using FigTree v1.4.4 (http:// tree.bio.ed.ac.uk/software/figtree/).

2.5. Selection pressure analysis

The complete coding sequences of SARS-CoV-2 Omicron VOC were screened for the presence of recombination breakpoints using the Genetic Algorithm for Recombination Detection (GARD) method, implemented in the HyPhy package's Datamonkey web server [41]. This screening is necessary because the inference based on the data with recombination breakpoints frequently yields more false positive sites [42,43]. In the cases of data with recombination breakpoints, the sequences were partitioned into the recombination blocks, and the selection pressure was estimated individually for each block. Then, using the Single Likelihood Ancestor Counting (SLAC) method, the site-specific selection pressure within SARS-CoV-2 Omicron variant was estimated as the ratio of nonsynonymous (dN) to synonymous (dS) nucleotide substitutions per site ($\omega = dN/dS$) [41].

We tested and compared the results of sites under selection pressure estimated by five different methods with their default parameters, namely SLAC [41], Fixed Effect Likelihood (FEL), Fast Unbiased Bayesian AppRoximation (FUBAR) [44], adaptive Branch-Site Random Effects Likelihood (aBSREL) [45,46], and Branch-Site Unrestricted Statistical Test for Episodic Diversification (BUSTED) [47], all available at the Datamonkey web server. Furthermore, the sites in each ORF of SARS-CoV-2 Omicron variant experiencing positive/diversifying and negative/purifying selection pressure were taken into account only when anticipated by at least two aforementioned methods.

3. Results and Discussion

3.1. Mutational scanning of SARS-CoV-2 Omicron reveals its independent emergence

The recently discovered Omicron VOC is distantly related to earlier documented VOCs (Alpha, Beta, Gamma, and Delta) and has a remarkably high number of mutations in the spike protein (Figure 1A). As a result of Omicron's continued genetic evolution, two sub-lineages (BA.1 and BA.2) were initially identified; BA.1 emerged as the dominant sub-lineage in the late 2021 and spread quickly throughout the world, but BA.2 quickly overtook BA.1 on a global scale (Figure 1B). The identification of a highly divergent Omicron VOC advocated the possibility that Omicron may have evolved in a cellular micro-environment completely different from other known VOCs. In comparison to the wild-type SARS-CoV-2 (WT), Omicron's receptor-binding domain (RBD) contains 15 mutations, ten of which are in the receptor-binding motif (RBM), which mediates binding to host cells via angiotensin-converting enzyme 2 (ACE2) receptor and to the majority of NAbs (Figure

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 October 2022 doi:10.20944/preprints202210.0404.v1

2). Four mutations in Omicron's RBM (N440K, S477N, T478K and N501Y), and one in RBD (G339D) have been demonstrated to improve binding affinity to human ACE2 [48,49].

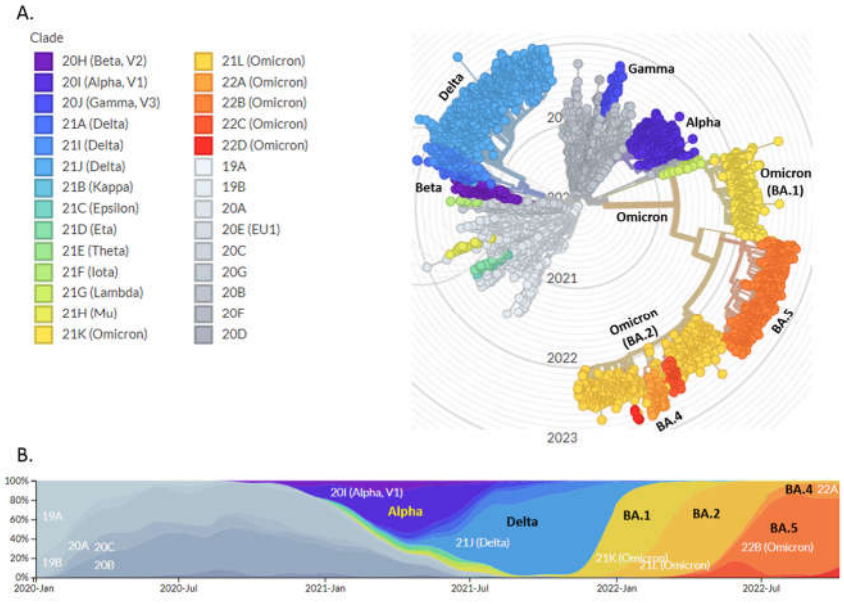

Figure 1. (A.) A time-resolved maximum likelihood phylogeny of a representative global subsample of 3004 SARS-CoV-2 complete genomes sampled between December 2019 and October 2022 from the GISAID. The variants of concern (VOCs) are denoted by different colour schemes. The figure was generated by the Nextstrain using the data from the GISAID, (B.) The global distribution frequencies of VOCs are denoted by different colour schemes over the timespan since SARS-CoV-2 emergence.

Furthermore, eleven RBD mutations (G339D, S371L, S373P, S375F, N440K, G446S, E484A, Q493R, G496S, Q498R, and Y505H) have been shown to be associated with escape from a wide range of different classes of NAbs [50-56]. These RBD mutations, together with N-terminal domain (NTD) mutations, particularly H69–V70del and G142–Y144del, favor Omicron's evasion from the majority of NAbs induced either by infections or vaccinations [15,57]. Notably, the mutation profile of Omicron BA.2's NTD (T19I, L24-P26del, A27S, G142D, and V213D) differs significantly from that of BA.1, although their functional roles remain unknown (Figure 2).

Some mutations in the Omicron' spike may also contribute in modulating the virus host spectrum. For example, acquisition of positively charged amino acids at 493 and 498 (Q493K and Q498H) has been shown to allow SARS-CoV-2 to infect mice via interacting with murine ACE2 [58,59]. These two mutations, Q493R and Q498R (both containing positively charged amino acids and found in Omicron's RBM), were acquired after 30 passages in mouse lung (GISAID accession number EPI_ISL_1666328) [60]. Furthermore, H655Y was selected during replication in the mink model, implying a role in modifying the host range [61]. As a result, Q493R, Q498R, and H655Y carried by Omicron' spike reflects its adaptation in mice and mink. These findings, together with a recent study demonstrating Omicron's ability to mediate the enhanced entry into cells expressing multiple animal species ACE2 [62], imply that, Omicron may have a broader host range and a greater proclivity to establish an animal reservoir for its family than previously known VOCs.

Intriguingly, of thirty mutations in the Omicron's spike protein, it shares only eight mutations with other known VOCs, including H69-V70del and P681H in Alpha, K417N in Beta, H655Y in Gamma, T19I and T478K in Delta, N501Y in Alpha, Beta, and Gamma,

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 October 2022 doi:10.20944/preprints202210.0404.v1

and D614G in Alpha, Beta, Gamma, and Delta. Omicron's unique mutation profile, combined with the amino acid substitutions pattern's low similarity to other known VOCs, opens the door for designating any of the previously reported VOCs or other variants as its most recent common ancestor. Furthermore, we noted distinct sub-lineage-specific mutations in the spike protein of BA.1 (A67V, T95I, G142-Y144del, Y145D, N211del, L212I, 214EPEins, S371L, G446S, T547K, N856K, and L981F) and BA.2 (T19I, L24-P26del, A27S, G142D, V213G, S371F, T376A, D405N, and R408S) possibly indicate their separate/independent emergence and evolution. Furthermore, regardless of the vaccination status, these sub-lineage-specific mutations may be associated with a higher susceptibility of infection by the BA.2 sub-lineage in comparison to the BA.1 sub-lineage [63].

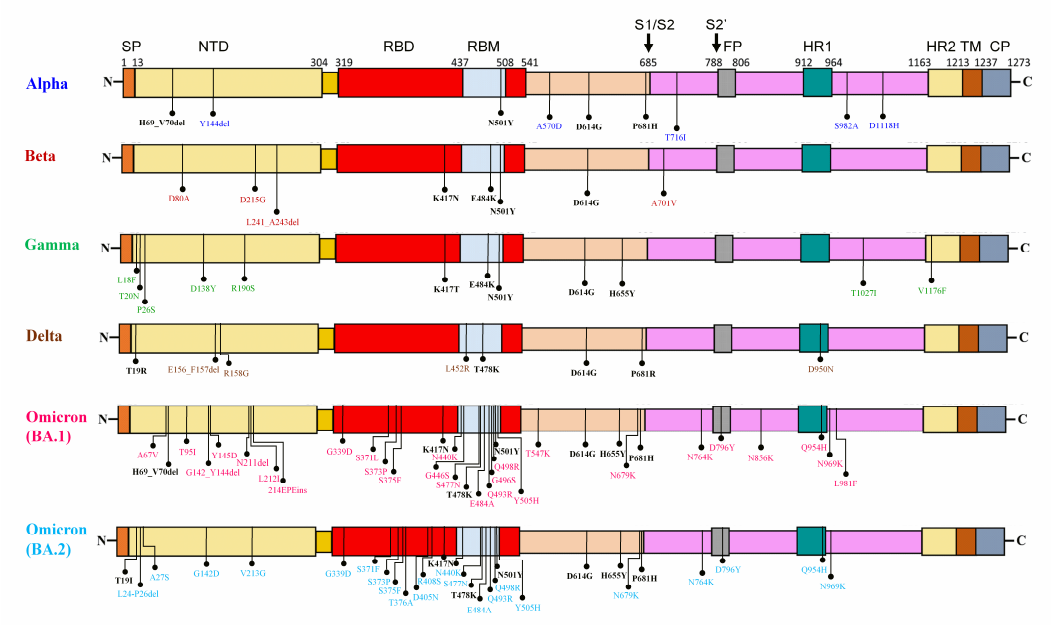

Figure 2. Mutational profiles of Variants of Concern's spike protein. Mutations that differed from each other VOCs are denoted by different VOCs-specific colours, while the shared mutations are represented by black colour. SP, Signal Peptide; NTD, N-terminal domain; RBD, Receptor-Binding Domain; RBM, Receptor-Binding Motif; SD1, subdomain 1; SD2, subdomain 2; FP, Fusion peptide; HR1, Heptad repeat region 1; HR2, Heptad repeat region 2; TM, Transmembrane domain; CP, Cytoplasmic domain.

Out of a cluster of three substitutions (H655Y, N679K, and P681H) found near the S1/S2 furin cleavage site of Omicron's spike protein, two substitutions (P681H in Alpha, H655Y in Gamma, and P681R in Delta) have been demonstrated to facilitate cleavage of the spike protein and increase viral fusogenicity with the host cells [61,64-66]. Additionally, the spike's fusion peptides of both Omicron's sub-lineages (BA.1 and BA.2) contained a D796Y substitution that was absent from the previously identified VOCs. The combination of these substitutions may enhance the fusogenicity and transmissibility of Omicron [61].

Other than the spike glycoprotein, Omicron has several mutations in other proteins. The BA.1 sub-lineage is distinct from the BA.2 sub-lineage in that the former had three substitutions (K38R, L1266I, and A1892T) and one deletion (S1265 del) in nsp3, but the latter did not. The rest of mutations are common in both the sub-lineages, including T492I in nsp4, P132H in nsp5, S106-G107-F108 deletion and I189V in nsp6, P323L in NS12, and I42V in nsp14. Nonetheless, little is known about their functional roles: a deletion in nsp6 (del105–107) for evasion of innate immunity [67], P323L in nsp12 (RNA-dependent RNA polymerase) for reduced binding affinity to remdesivir [68], and two mutations (R203K and G204R) in nucleocapsid for enhanced infectivity [69].

 doi:10.20944/preprints202210.0404.v1

3.2. Recombination analysis

Recombination is a fundamental mechanism for generating diversity among positive-sense RNA viruses, including SARS-CoV-2, and is an important tool for understanding the evolutionary history of viruses. Furthermore, the Bayesian molecular dating analyses on the dataset having evidence of recombination can result in the biased phylogenetic and phylodynamic inferences [70,71]. As a result, making of the dataset free from the recombination signals is a crucial step in deriving the molecular clock phylogenetics inferences.

We individually screened the sequences of the Omicron BA.1 and BA.2 sub-lineages for recombination signals using the RDP v4.101, which implements nine distinct algorithms to locate evidences of recombination signals. In the SARS-CoV-2 Omicron BA.1 sub-lineage dataset, we found a total of four recombination signals that were recognized by at least two different algorithms (Table S2). Three of these, however, lacked a high level of evidence. We found moderate evidence for only one recombination signal, which was identified by Chimaera ($P = 0.0025$), 3Seq ($P = 0.0005$), and Maxchi ($P < 0.0001$) in the NTD encoding region of spike protein. The breakpoint positions for this moderate recombinant signal were 21613 for the 5′ breakpoint and 27265 for the 3′ breakpoint. However, the recombination analyses could not identify the recombination signals in the Omicron BA.2 lineage dataset. The identified recombinant sequence, hCoV-19/env/Austria/CeMM21831/2022|EPI_ISL_9011265, could have 78 minor and 14 major parental sequences. Lastly, we identified and removed all the recombinant sequences projected to convey even a low level of evidence from the dataset before conducting the Bayesian molecular dating analyses.

3.3. Bayesian molecular dating analyses of Omicron VOC

After screening 88 distinct nucleotide substitution models for the SARS-CoV-2 Omicron sub-lineages (BA.1 and BA.2) datasets, General Time Reversible (GTR+F+I) and Tamura-Nei (TN+F+I) models were found to be the best nucleotide substitution models for BA.1 and BA.2 sub-lineages, respectively (Table S1). The ML trees were generated using these best fitting nucleotide substitution models. Using TempEst v1.5.3, root-to-tip regression analysis was performed on the ML trees generated separately for the Omicron BA.1 and BA.2 sub-lineages to assess the temporal molecular evolutionary signals. The coefficient of the determinant, R2, which measures the clock-likeness of the sequences, and the correlation of coefficient (r) was low for BA.1 ($r = 0.371$, and $R2 = 0.137$) and BA.2 ($r = 0.166$, and $R2 = 0.027$) dataset.

To improve the temporal signals in our dataset, we first identified and removed any outliers (often caused by sequencing errors or incorrect labeling) that did not fit to a root-to-tip regression. Subsequently, while maintaining the sequence heterogeneity, we generated three datasets by randomly picking up unique Omicron sequences from each country representing; (i) one sequence from each sampling date, (ii) two sequences from each sampling date, and (iii) three sequences from each sampling date. By doing this, temporal signals for both Omicron sub-lineages BA.1 ($r = 0.454$, and $R2 = 0.206$) and BA.2 ($r = 0.549$, and $R2 = 0.302$) improved significantly in the second set of data (Figure 3, Table S3).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 October 2022 doi:10.20944/preprints202210.0404.v1

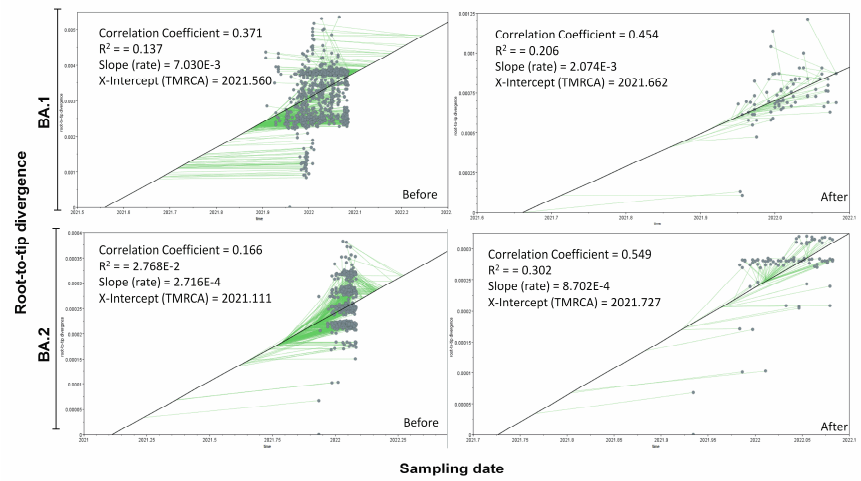

Figure 3. Root-to-tip regression analyses on the maximum likelihood (ML) trees generated for original dataset comprising Omicron BA.1 and BA.2 sub-lineages (Before) and dataset after removing the outliers (After).

Next, Bayesian molecular dating analyses was performed on the second set of datasets containing 161 dated non-recombinant nucleotide sequences for the complete coding sequences of Omicron. We compared the prior, posterior, and likelihood distributions of each of the eight tree priors-clocks combinations in order to determine the best model-fit for the Omicron sub-lineages (Figure 4A-F). The estimated tMRCA dates and evolutionary rates of BA.1 sub-lineage was relatively comparable across all the tree priors (e.g., constant population size, exponential population, Bayesian skyline, extended Bayesian skyline), but varied greatly depending on the clock model used. Comparatively, the estimated tMRCA dates and evolutionary rates of BA.2 sub-lineage was very similar depending on the clock model used, but varied across different tree priors (Table S4).

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 October 2022 doi:10.20944/preprints202210.0404.v1

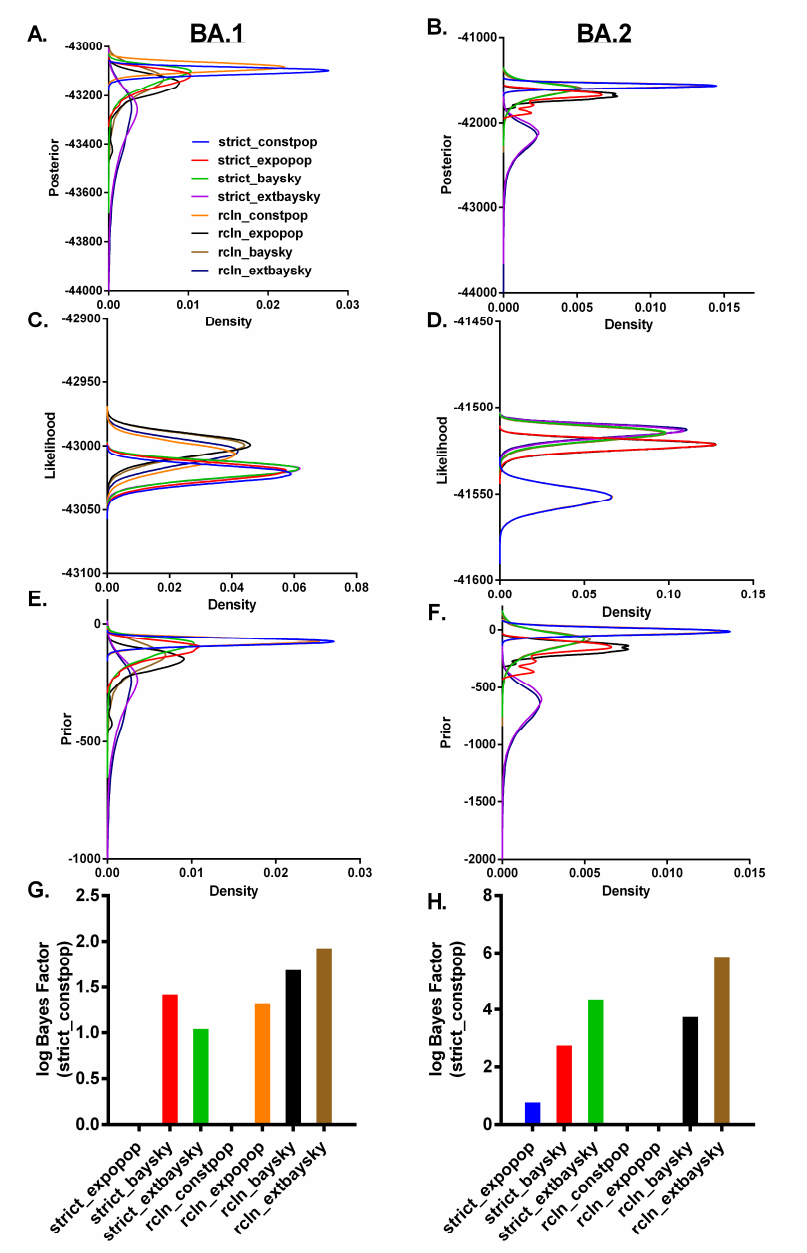

Figure 4. Bayesian hypothesis testing using the log Bayes factor. Posterior (A, B), Likelihood (C, D) and Prior (E, F) estimates for comparative evaluation of eight combinations of tree priors and clock models implemented in BEAST v2.6.7 are shown for BA.1, and BA.2 sub-lineages, respectively. (G, H) Bayesian hypothesis testing using the log Bayes factor. The best fit combination of tree priors and clock models was constant population coalescent tree prior with strict clock for both the Omicron sub-lineages, BA.1 and BA.2.

The Constant Population coalescent tree prior with strict clock is the best fit to both the Omicron sub-lineages, according to Bayesian hypothesis testing using the log Bayes factor (Figure 4G-H). A time-scaled maximum-clade-credibility tree showed that all of the omicron sequences could be separated into two distinct sub-lineages, BA.1 (n = 75) and BA.2 (n = 86). The most recent common ancestor (tMRCA) for BA.1 and BA.2 sub-lineages sequences was estimated to be 18 September 2021 (95% highest posterior density (HPD) 04 August – 22 October 2021), and 03 November 2021 (95% HPD 26 September – 28 November 2021), respectively (Figure 5). The substitution rates of BA.1 and BA.2 were estimated to be 1.435×10^{-3} (95% HPD = 1.021×10^{-3} – 1.869×10^{-3}) substitution/site/year and

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 October 2022 doi:10.20944/preprints202210.0404.v1

1.074×10^{-3} (95% HPD = 6.444×10^{-4} – 1.586×10^{-3}) substitution/site/year, respectively, which in line with several previous studies that estimated substitution rates of SARS-CoV-2 [9,72-74].

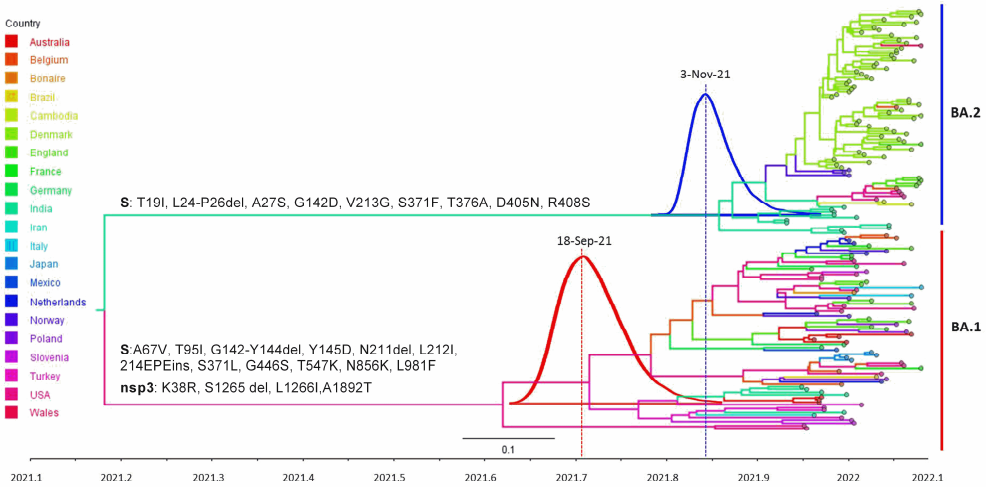

Figure 5. Time-resolved maximum clade credibility phylogeny of the Omicron BA.1 and BA. 2 sub-lineages (n = 161, sampled between November 2021 and January 2022). Mutations that characterize Omicron sub-lineages are indicated on the respective branches. The posterior distribution of the tMRCA for each sub-lineage is also shown. The countries as a trait for each sample is denoted by distinct colours.

In comparison to our dataset tMRCA for BA.1 sub-lineage, the estimated tMRCA of the South African BA.1 sub-lineage sequence was found to be early-October 2021 (9 October 2021, 95% HPD 30 September–20 October 2021) [9]. This discrepancy is likely due to differences in dataset's geographical heterogeneity. The tMRCA estimate for the BA.2 sub-lineage is in perfect accord with other studies, such as the tMRCA of BA.2 sub-lineage on 6 November 2021 (95% HPD = 9 October 2021 to 29 November 2021) [19], and mid-November for Philippines BA.2 lineage sequences (18 November 2021; 95 % HPD = 6–28 November 2021) [75]. Because the BA.1 and BA.2 sub-lineages exhibit different tMRCA and have distinct mutational profiles, it is possible that these two sub-lineages of Omicron VOC might have originated and evolved independently.

The evolutionary history of Omicron is presently governed by three hypotheses. The first hypothesis is that Omicron might have spread silently in a geographical region with limited surveillance and sequencing facility [76,77]. Second, Omicron might have evolved in an immunocompromised patient, allowing long-term sustained evolution and adaptation of the virus [76,78]. Third, the Omicron might have accumulated mutations in a non-human host before jumping to humans [78,79]. Presently, the second hypothesis seems to be more plausible regarding the evolutionary origin of Omicron [79]. Nonetheless, in the event of the emergence of multiple new mutations in the Omicron's spike protein, which are quite distinct in the BA.1 and BA.2 sub-lineages, as well as their estimated separate most recent common ancestor, it may be more plausible to conclude that a combination of RBD- and NTD -directed different classes of antibody therapeutics at the sub-optimal doses in COVID-19 patients or optimal doses in an immunocompromised patient or waned vaccine-induced immunity may have provided a conducive environment to accumulate multiple mutations in Omicron's spike protein. However, the role of intermediate hosts, particularly rodents, in Omicron transmission and evolution, followed by reverse-zoonosis, should not be neglected.

 doi:10.20944/preprints202210.0404.v1

Furthermore, a per-day exponential growth rate of 0.151 (95% HPD 0.047–0.243) for BA.1, and 0.155 (95% HPD 0.007–0.392) for BA.2 was estimated, indicating that BA.2 has a slight growth advantage over BA.1, which is in agreement with a previous study [19]. The growth advantage of BA.2 could be due to its inherent enhanced transmissibility. This observation is in concordance with a recent in vitro study that found BA.2 to form the larger syncytia and be more contagious in human nasal epithelial cells than BA.1 [80]. Nonetheless, the exponential growth rate is also influenced by the immune status of patients, social distancing, housing and ventilation, and the superspreading events; thus, these factors may have a significant impact on the exponential growth rate of the viruses.

3.4. Selection analysis

The rapid evolution of RNA viruses, particularly SARS-CoV-2, is governed by increasing and persistent selection pressure, leading in the creation of viruses with altered genetic and phenotypic characteristics [81,82]. These evolutionary changes are the result of two opposing forces; positive selection (which generates various genetic changes beneficial for virus enhanced fitness to its host), and negative selection (that tries to maintain fitness without producing any advantageous changes). Therefore, positive selection may have accelerated the accumulation of multiple mutations in the Omicron BA.1 and BA.2 sub-lineages. To test for evidence of selection (both positive/diversifying and negative/purifying selection) in the Omicron's evolution, we employed selection pressure estimation pipeline comprising five methods: SLAC, FEL, FUBAR, aBSREL, and BUSTED at the Datamonkey web server of the HyPhy package. We filtered out sequences that do not have Omicron signature mutational profiles and then extracted individual protein coding regions (11 ORFs) from unique BA.1 (n = 689) and BA.2 (n = 948) sub-lineages to estimate site-specific selection pressure. To improve the robustness of site-selection and reduce the false positive rates, only sites predicted by at least two of the aforementioned methods were considered.

Among the 11 ORFs, three ORFs (ORF3a, ORF7a, and ORF7b) had experienced a strong selection (ratio of non-synonymous to synonymous substitutions is more than 1.5). There are seven positively selected sites in ORF1ab, 13 in spike protein and 1 in Nucleocapsid protein of the Omicron BA.1 sub-lineage, and only one in Membrane protein of BA.2 sub-lineage, indicating that multiple codon sites drive the genetic diversity in BA.1 sub-lineage (Table 1). Furthermore, of the spike's 13 sites in the BA.1 sub-lineage, eight sites (339, 371, 375, 440, 446, 484, 493, and 505) that showed evidence of diversifying or positive selection, are associated with escape from different classes of NAbs [50-56], implying that these sites are still evolving in order to modify BA.1's neutralization profile. Importantly, four sites in S (346, 452, 554, and 1260), one each in N (215) and nsp6 (3646) that showed positive selection signals are not Omicron-defining mutations, and these sites could have been carried by its most recent common ancestor. However, three sites carrying the positive selection signals converge on mutations found in previously identified SARS-CoV-2 VOCs (S:452, N:215, and nsp6:3646 in delta; S:346 in Mu VOI). These observations support the notion that Omicron might have originated and evolved independently.

Preprints (www.preprints.org) | NOT PEER-REVIEWED | Posted: 26 October 2022 doi:10.20944/preprints202210.0404.v1

Table 1. Amino acid sites under diversifying or purifying selection for each protein encoded regions of Omicron BA.1 and BA.2 sub-lineages .

	BA.1			BA.2		
ORFs	**dN/dS**	**Diversifying selection**	**Purifying selection**	**dN/dS**	**Diversifying selection**	**Purifying selection**
ORF1ab	0.580	nsp3 (2710); nsp4 (3255); nsp5 (3395); nsp6 (3646, 3758); nsp12 (5063); nsp14 (5967)	nsp2 (341, 735); nsp3 (1120, 1707, 1750, 1903, 2676); nsp4 (2907); nsp5 (3290, 3458); nsp6 (3689); nsp10 (4310); nsp12 (4992); nsp13 (5444, 5541)	0.525	-	nsp2 (440); nsp3 (924, 1707, 2470, 2551); nsp4 (3100, 3245); nsp5 (3271, 3311); nsp13 (5616, 5746); nsp15 (6566); nsp16 (6819)
S	0.981	RBD (339, 346, 371, 375, 440, 446, 452, 484, 493, 505); SD1/SD2 (554); S2 (796, 1260)	SP (11); SD1/SD2 (543); S2 (1146)	0.653	-	NTD (296); RBD (336, 410)
ORF3a	2.040	-	-	1.266	-	-
E	0.401	-	-	0.341	-	67, 68
M	0.201	-	135	0.477	3	-
ORF6	0.259	-	-	0.597	-	-
ORF7a	1.52	-	-	1.91	-	-
ORF7b	1.67	-	-	1.37	-	-
ORF8	0.563	-	-	0.986	-	-
N	0.812	215	73, 324	0.880	-	329
ORF10	0.319	-	-	0.970	-	-

4. Conclusions

The findings of this study, which investigated the early global evolutionary dynamics of the recently identified, highly divergent SARS-CoV-2 Omicron VOC sampled between November 2021 and January 2022, suggest that the Omicron BA.1 and BA.2 sub-lineages originated independently and evolved over time. This study also advocates the continued diversifying selection that may alter the neutralization profile of BA.1. Numerous mutations, particularly in spike's NTD and nsp3, whose functions are still unclear, warrant functional characterization in order to understand their contributions to differential viral transmissibility and diminished efficacy of therapeutics and vaccines. The currently available evidence supports the idea that the origin and evolution of the Omicron may have been influenced by the long-term persistence of infections in human and/or animal hosts (mainly as an intermediate host for reverse-zoonosis). Finally, this study emphasizes the significance of ongoing global genomic surveillance, Bayesian molecular dating analyses, and mutational profiling in understanding the virus's evolutionary dynamics and, as a result, mitigating the impact of emerging variants on public health.

Supplementary Materials: Table S1: Identification of best fitting nucleotide substitution model based on the Bayesian Information Criterion (BIC) using the ModelFinder; Table S2: The recombination signals detected in the Omicron BA.1 sub-lineage; Table S3: Temporal signals in three datasets (of BA.1 and BA.2 sub-lineages) generated by randomly picking up unique Omicron sequences from each country representing; (i) one sequence from each sampling date (1SEQ), (ii) two sequences from each sampling date (2SEQ), and (iii) three sequences from each sampling date (3SEQ); Table S4: A comparison of the estimated tMRCA dates and evolutionary rates of Omicron's BA.1 and BA.2 sub-lineages across all the tree priors and clock models combinations.

Author Contributions: Conceptualization, N.K.; Data curation, N.K. and R.K.; Formal analysis, N.K., R.K. and A.S.; Methodology, N.K. and R.K.; Software, N.K., R.K. and K.Y.J.Z; Validation, N.K. and R.K.; Writing—original draft, N.K. and R.K.; Writing—review and editing, N.K., R.K., A.S., K.Y.J.Z., V.N.U., U.S., S.B. and A.S. All authors have read and agreed to the published version of the manuscript.

Funding: This study was funded in part by the Indian Council of Agricultural Research (ICAR) - National Agricultural Science Fund (NASF/ABA-8028/2020-21 to S.B. and N.K.)

Acknowledgments: The authors acknowledge the Advanced Center for Computing and Communication (ACCC) of the Institute of Physical and Chemical Research (RIKEN) for computing resources on the Hokusai BigWaterfall supercomputer, and Biotech Research Center of the Technology Innovation Institute (TII) for support and resources. The authors also thanks Prof. Thomas Launey, TII for critical reading and improvements of the manuscript. Finally, all of the authors thank and express gratitude to their respective Institutes.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. GISIAD. Clade and lineage nomenclature aids in genomic epidemiology studies of active hCoV-19 viruses. Available online: https://go.nature.com/3pgSIt6 (accessed on 21/10/2022).
2. Rambaut, A.; Holmes, E.C.; O'Toole, Á.; Hill, V.; McCrone, J.T.; Ruis, C.; du Plessis, L.; Pybus, O.G. A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. *Nat Microbiol* **2020**, *5*, 1403-1407, doi:10.1038/s41564-020-0770-5.
3. Trevor, B.; Hodcroft, E.B.; Neher, R.A. Updated Nextstrain SARS-CoV-2 clade naming strategy. Available online: https://nextstrain.org/blog/2021-01-06-updated-SARS-CoV-2-clade-naming (accessed on 10/21/2022).
4. WHO. Tracking SARS-CoV-2 variants. Available online: https://www.who.int/activities/tracking-SARS-CoV-2-variants (accessed on 01/10/2022).
5. Alai, S.; Gujar, N.; Joshi, M.; Gautam, M.; Gairola, S. Pan-India novel coronavirus SARS-CoV-2 genomics and global diversity analysis in spike protein. *Heliyon* **2021**, *7*, e06564, doi:10.1016/j.heliyon.2021.e06564.
6. Faria, N.R.; Mellan, T.A.; Whittaker, C.; Claro, I.M.; Candido, D.D.S.; Mishra, S.; Crispim, M.A.E.; Sales, F.C.; Hawryluk, I.; McCrone, J.T.; et al. Genomics and epidemiology of a novel SARS-CoV-2 lineage in Manaus, Brazil. *medRxiv* **2021**, doi:10.1101/2021.02.26.21252554.
7. Tegally, H.; Wilkinson, E.; Giovanetti, M.; Iranzadeh, A.; Fonseca, V.; Giandhari, J.; Doolabh, D.; Pillay, S.; San, E.J.; Msomi, N.; et al. Detection of a SARS-CoV-2 variant of concern in South Africa. *Nature* **2021**, *592*, 438-443, doi:10.1038/s41586-021-03402-9.
8. Dhar, M.S.; Marwal, R.; Vs, R.; Ponnusamy, K.; Jolly, B.; Bhoyar, R.C.; Sardana, V.; Naushin, S.; Rophina, M.; Mellan, T.A.; et al. Genomic characterization and epidemiology of an emerging SARS-CoV-2 variant in Delhi, India. *Science* **2021**, *374*, 995-999, doi:10.1126/science.abj9932.
9. Viana, R.; Moyo, S.; Amoako, D.G.; Tegally, H.; Scheepers, C.; Althaus, C.L.; Anyaneji, U.J.; Bester, P.A.; Boni, M.F.; Chand, M.; et al. Rapid epidemic expansion of the SARS-CoV-2 Omicron variant in southern Africa. *Nature* **2022**, *603*, 679-686, doi:10.1038/s41586-022-04411-y.
10. Kannan, S.R.; Spratt, A.N.; Sharma, K.; Chand, H.S.; Byrareddy, S.N.; Singh, K. Omicron SARS-CoV-2 variant: Unique features and their impact on pre-existing antibodies. *J Autoimmun* **2022**, *126*, 102779, doi:10.1016/j.jaut.2021.102779.
11. Ulloa, A.C.; Buchan, S.A.; Daneman, N.; Brown, K.A. Estimates of SARS-CoV-2 Omicron Variant Severity in Ontario, Canada. *Jama* **2022**, *327*, 1286-1288, doi:10.1001/jama.2022.2274.
12. Pulliam, J.R.C.; van Schalkwyk, C.; Govender, N.; von Gottberg, A.; Cohen, C.; Groome, M.J.; Dushoff, J.; Mlisana, K.; Moultrie, H. Increased risk of SARS-CoV-2 reinfection associated with emergence of Omicron in South Africa. *Science* **2022**, *376*, eabn4947, doi:10.1126/science.abn4947.
13. Augusto, G.; Mohsen, M.O.; Zinkhan, S.; Liu, X.; Vogel, M.; Bachmann, M.F. In vitro data suggest that Indian delta variant B.1.617 of SARS-CoV-2 escapes neutralization by both receptor affinity and immune evasion. *Allergy* **2022**, *77*, 111-117, doi:10.1111/all.15065.
14. Han, P.; Su, C.; Zhang, Y.; Bai, C.; Zheng, A.; Qiao, C.; Wang, Q.; Niu, S.; Chen, Q.; Zhang, Y.; et al. Molecular insights into receptor binding of recent emerging SARS-CoV-2 variants. *Nat Commun* **2021**, *12*, 6103, doi:10.1038/s41467-021-26401-w.
15. Cele, S.; Jackson, L.; Khoury, D.S.; Khan, K.; Moyo-Gwete, T.; Tegally, H.; San, J.E.; Cromer, D.; Scheepers, C.; Amoako, D.G.; et al. Omicron extensively but incompletely escapes Pfizer BNT162b2 neutralization. *Nature* **2022**, *602*, 654-656, doi:10.1038/s41586-021-04387-1.
16. Dejnirattisai, W.; Shaw, R.H.; Supasa, P.; Liu, C.; Stuart, A.S.; Pollard, A.J.; Liu, X.; Lambe, T.; Crook, D.; Stuart, D.I.; et al. Reduced neutralisation of SARS-CoV-2 omicron B.1.1.529 variant by post-immunisation serum. *Lancet* **2022**, *399*, 234-236, doi:10.1016/s0140-6736(21)02844-0.
17. Khoury, D.S.; Steain, M.; Triccas, J.A.; Sigal, A.; Davenport, M.P.; Cromer, D. Analysis: A meta-analysis of Early Results to predict Vaccine efficacy against Omicron. *medRxiv* **2021**, 2021.2012.2013.21267748, doi:10.1101/2021.12.13.21267748.
18. Kumar, N.; Kaushik, R.; Zhang, K.Y.J.; Uversky, V.N.; Srivastava, P.; Sahu, U.; Sood, R.; Bhatia, S. A novel consensus-based computational pipeline for rapid screening of antibody therapeutics for efficacy against SARS-CoV-2 variants of concern including omicron variant. *bioRxiv* **2022**, 2022.2002.2011.480177, doi:10.1101/2022.02.11.480177.

19. Tegally, H.; Moir, M.; Everatt, J.; Giovanetti, M.; Scheepers, C.; Wilkinson, E.; Subramoney, K.; Makatini, Z.; Moyo, S.; Amoako, D.G.; et al. Emergence of SARS-CoV-2 Omicron lineages BA.4 and BA.5 in South Africa. *Nat Med* **2022**, *28*, 1785-1790, doi:10.1038/s41591-022-01911-2.
20. Elbe, S.; Buckland-Merrett, G. Data, disease and diplomacy: GISAID's innovative contribution to global health. *Glob Chall* **2017**, *1*, 33-46, doi:10.1002/gch2.1018.
21. Katoh, K.; Standley, D.M. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol* **2013**, *30*, 772-780, doi:10.1093/molbev/mst010.
22. Hall, T.A. BIOEDIT: A USER-FRIENDLY BIOLOGICAL SEQUENCE ALIGNMENT EDITOR AND ANALYSIS PROGRAM FOR WINDOWS 95/98/ NT. 1999.
23. Martin, D.P.; Varsani, A.; Roumagnac, P.; Botha, G.; Maslamoney, S.; Schwab, T.; Kelz, Z.; Kumar, V.; Murrell, B. RDP5: a computer program for analyzing recombination in, and removing signals of recombination from, nucleotide sequence datasets. *Virus Evol* **2021**, *7*, veaa087, doi:10.1093/ve/veaa087.
24. Kalyaanamoorthy, S.; Minh, B.Q.; Wong, T.K.F.; von Haeseler, A.; Jermiin, L.S. ModelFinder: fast model selection for accurate phylogenetic estimates. *Nat Methods* **2017**, *14*, 587-589, doi:10.1038/nmeth.4285.
25. Nguyen, L.T.; Schmidt, H.A.; von Haeseler, A.; Minh, B.Q. IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Mol Biol Evol* **2015**, *32*, 268-274, doi:10.1093/molbev/msu300.
26. Rambaut, A.; Lam, T.T.; Max Carvalho, L.; Pybus, O.G. Exploring the temporal structure of heterochronous sequences using TempEst (formerly Path-O-Gen). *Virus Evolution* **2016**, *2*, vew007, doi:10.1093/ve/vew007.
27. Drummond, A.J.; Nicholls, G.K.; Rodrigo, A.G.; Solomon, W. Estimating mutation parameters, population history and genealogy simultaneously from temporally spaced sequence data. *Genetics* **2002**, *161*, 1307-1320, doi:10.1093/genetics/161.3.1307.
28. Bouckaert, R.; Vaughan, T.G.; Barido-Sottani, J.; Duchêne, S.; Fourment, M.; Gavryushkina, A.; Heled, J.; Jones, G.; Kühnert, D.; De Maio, N.; et al. BEAST 2.5: An advanced software platform for Bayesian evolutionary analysis. *PLoS Comput Biol* **2019**, *15*, e1006650, doi:10.1371/journal.pcbi.1006650.
29. Kingman, J.F.C. The coalescent. *Stochastic Processes and their Applications* **1982**, *13*, 235-248, doi:https://doi.org/10.1016/0304-4149(82)90011-4.
30. Griffiths, R.C.; Tavaré, S. Sampling theory for neutral alleles in a varying environment. *Philos Trans R Soc Lond B Biol Sci* **1994**, *344*, 403-410, doi:10.1098/rstb.1994.0079.
31. Drummond, A.J.; Rambaut, A.; Shapiro, B.; Pybus, O.G. Bayesian coalescent inference of past population dynamics from molecular sequences. *Mol Biol Evol* **2005**, *22*, 1185-1192, doi:10.1093/molbev/msi103.
32. Heled, J.; Drummond, A.J. Bayesian inference of population size history from multiple loci. *BMC Evol Biol* **2008**, *8*, 289, doi:10.1186/1471-2148-8-289.
33. Drummond, A.J.; Ho, S.Y.; Phillips, M.J.; Rambaut, A. Relaxed phylogenetics and dating with confidence. *PLoS Biol* **2006**, *4*, e88, doi:10.1371/journal.pbio.0040088.
34. Sinsheimer, J.S.; Lake, J.A.; Little, R.J. Bayesian hypothesis testing of four-taxon topologies using molecular sequence data. *Biometrics* **1996**, *52*, 193-210.
35. Kass, R.E.; Raftery, A.E. Bayes Factors. *Journal of the American Statistical Association* **1995**, *90*, 773-795, doi:10.1080/01621459.1995.10476572.
36. Lartillot, N.; Philippe, H. Computing Bayes factors using thermodynamic integration. *Syst Biol* **2006**, *55*, 195-207, doi:10.1080/10635150500433722.
37. Xie, W.; Lewis, P.O.; Fan, Y.; Kuo, L.; Chen, M.H. Improving marginal likelihood estimation for Bayesian phylogenetic model selection. *Syst Biol* **2011**, *60*, 150-160, doi:10.1093/sysbio/syq085.
38. Gelman, A.; Meng, X.L. Simulating Normalizing Constants: From Importance Sampling to Bridge Sampling to Path Sampling. *Statistical Science* **1998**, *13*, 22.
39. Rambaut, A.; Drummond, A.J.; Xie, D.; Baele, G.; Suchard, M.A. Posterior Summarization in Bayesian Phylogenetics Using Tracer 1.7. *Syst Biol* **2018**, *67*, 901-904, doi:10.1093/sysbio/syy032.
40. Drummond, A.J.; Suchard, M.A.; Xie, D.; Rambaut, A. Bayesian phylogenetics with BEAUti and the BEAST 1.7. *Mol Biol Evol* **2012**, *29*, 1969-1973, doi:10.1093/molbev/mss075.
41. Kosakovsky Pond, S.L.; Frost, S.D.W. Not So Different After All: A Comparison of Methods for Detecting Amino Acid Sites Under Selection. *Molecular Biology and Evolution* **2005**, *22*, 1208-1222, doi:10.1093/molbev/msi105.
42. Kosakovsky Pond, S.L.; Posada, D.; Gravenor, M.B.; Woelk, C.H.; Frost, S.D. Automated phylogenetic detection of recombination using a genetic algorithm. *Mol Biol Evol* **2006**, *23*, 1891-1901, doi:10.1093/molbev/msl051.
43. Sood, R.; Kumar, N.; Gokhe, S.S.; Pateriya, A.K.; Bhat, S.; Bhatia, S.; Panickan, S.; Mishra, A.; Murugkar, H.V.; Dixit, R.; et al. Identification and molecular characterization of H9N2 viruses carrying multiple mammalian adaptation markers in resident birds in central-western wetlands in India. *Infect Genet Evol* **2021**, *94*, 105005, doi:10.1016/j.meegid.2021.105005.
44. Murrell, B.; Moola, S.; Mabona, A.; Weighill, T.; Sheward, D.; Kosakovsky Pond, S.L.; Scheffler, K. FUBAR: A Fast, Unconstrained Bayesian AppRoximation for Inferring Selection. *Molecular Biology and Evolution* **2013**, *30*, 1196-1205, doi:10.1093/molbev/mst030.
45. Kosakovsky Pond, S.L.; Murrell, B.; Fourment, M.; Frost, S.D.W.; Delport, W.; Scheffler, K. A Random Effects Branch-Site Model for Detecting Episodic Diversifying Selection. *Molecular Biology and Evolution* **2011**, *28*, 3033-3043, doi:10.1093/molbev/msr125.

46. Smith, M.D.; Wertheim, J.O.; Weaver, S.; Murrell, B.; Scheffler, K.; Kosakovsky Pond, S.L. Less is more: an adaptive branch-site random effects model for efficient detection of episodic diversifying selection. *Mol Biol Evol* **2015**, *32*, 1342-1353, doi:10.1093/molbev/msv022.
47. Murrell, B.; Weaver, S.; Smith, M.D.; Wertheim, J.O.; Murrell, S.; Aylward, A.; Eren, K.; Pollner, T.; Martin, D.P.; Smith, D.M.; et al. Gene-wide identification of episodic selection. *Mol Biol Evol* **2015**, *32*, 1365-1371, doi:10.1093/molbev/msv035.
48. Kaushik, R.; Kumar, N.; Zhang, K.Y.J.; Srivastava, P.; Bhatia, S.; Malik, Y.S. A novel structure-based approach for identification of vertebrate susceptibility to SARS-CoV-2: Implications for future surveillance programmes. *Environ Res* **2022**, *212*, 113303, doi:10.1016/j.envres.2022.113303.
49. Starr, T.N.; Greaney, A.J.; Hilton, S.K.; Ellis, D.; Crawford, K.H.D.; Dingens, A.S.; Navarro, M.J.; Bowen, J.E.; Tortorici, M.A.; Walls, A.C.; et al. Deep Mutational Scanning of SARS-CoV-2 Receptor Binding Domain Reveals Constraints on Folding and ACE2 Binding. *Cell* **2020**, *182*, 1295-1310.e1220, doi:10.1016/j.cell.2020.08.012.
50. Cao, Y.; Wang, J.; Jian, F.; Xiao, T.; Song, W.; Yisimayi, A.; Huang, W.; Li, Q.; Wang, P.; An, R.; et al. Omicron escapes the majority of existing SARS-CoV-2 neutralizing antibodies. *Nature* **2022**, *602*, 657-663, doi:10.1038/s41586-021-04385-3.
51. Engelhart, E.; Lopez, R.; Emerson, R.; Lin, C.; Shikany, C.; Guion, D.; Kelley, M.; Younger, D. Massively multiplexed affinity characterization of therapeutic antibodies against SARS-CoV-2 variants. *Antib Ther* **2022**, *5*, 130-137, doi:10.1093/abt/tbac011.
52. Focosi, D.; Novazzi, F.; Genoni, A.; Dentali, F.; Gasperina, D.D.; Baj, A.; Maggi, F. Emergence of SARS-COV-2 Spike Protein Escape Mutation Q493R after Treatment for COVID-19. *Emerg Infect Dis* **2021**, *27*, 2728-2731, doi:10.3201/eid2710.211538.
53. Liu, L.; Iketani, S.; Guo, Y.; Chan, J.F.; Wang, M.; Liu, L.; Luo, Y.; Chu, H.; Huang, Y.; Nair, M.S.; et al. Striking antibody evasion manifested by the Omicron variant of SARS-CoV-2. *Nature* **2022**, *602*, 676-681, doi:10.1038/s41586-021-04388-0.
54. Long, S.W.; Olsen, R.J.; Christensen, P.A.; Bernard, D.W.; Davis, J.J.; Shukla, M.; Nguyen, M.; Saavedra, M.O.; Yerramilli, P.; Pruitt, L.; et al. Molecular Architecture of Early Dissemination and Massive Second Wave of the SARS-CoV-2 Virus in a Major Metropolitan Area. *mBio* **2020**, *11*, doi:10.1128/mBio.02707-20.
55. Starr, T.N.; Greaney, A.J.; Dingens, A.S.; Bloom, J.D. Complete map of SARS-CoV-2 RBD mutations that escape the monoclonal antibody LY-CoV555 and its cocktail with LY-CoV016. *Cell Rep Med* **2021**, *2*, 100255, doi:10.1016/j.xcrm.2021.100255.
56. Tandel, D.; Gupta, D.; Sah, V.; Harinivas Harshan, K. N440K variant of SARS-CoV-2 has Higher Infectious Fitness. *bioRxiv* **2021**, 2021.2004.2030.441434, doi:10.1101/2021.04.30.441434.
57. Planas, D.; Saunders, N.; Maes, P.; Guivel-Benhassine, F.; Planchais, C.; Buchrieser, J.; Bolland, W.H.; Porrot, F.; Staropoli, I.; Lemoine, F.; et al. Considerable escape of SARS-CoV-2 Omicron to antibody neutralization. *Nature* **2022**, *602*, 671-675, doi:10.1038/s41586-021-04389-z.
58. Huang, K.; Zhang, Y.; Hui, X.; Zhao, Y.; Gong, W.; Wang, T.; Zhang, S.; Yang, Y.; Deng, F.; Zhang, Q.; et al. Q493K and Q498H substitutions in Spike promote adaptation of SARS-CoV-2 in mice. *EBioMedicine* **2021**, *67*, 103381, doi:10.1016/j.ebiom.2021.103381.
59. Liu, K.; Pan, X.; Li, L.; Yu, F.; Zheng, A.; Du, P.; Han, P.; Meng, Y.; Zhang, Y.; Wu, L.; et al. Binding and molecular basis of the bat coronavirus RaTG13 virus to ACE2 in humans and other species. *Cell* **2021**, *184*, 3438-3451.e3410, doi:10.1016/j.cell.2021.05.031.
60. Shu, Y.; McCauley, J. GISAID: Global initiative on sharing all influenza data - from vision to reality. *Euro Surveill* **2017**, *22*, doi:10.2807/1560-7917.Es.2017.22.13.30494.
61. Escalera, A.; Gonzalez-Reiche, A.S.; Aslam, S.; Mena, I.; Laporte, M.; Pearl, R.L.; Fossati, A.; Rathnasinghe, R.; Alshammary, H.; van de Guchte, A.; et al. Mutations in SARS-CoV-2 variants of concern link to increased spike cleavage and virus transmission. *Cell Host Microbe* **2022**, *30*, 373-387.e377, doi:10.1016/j.chom.2022.01.006.
62. Peacock, T.P.; Brown, J.C.; Zhou, J.; Thakur, N.; Newman, J.; Kugathasan, R.; Sukhova, K.; Kaforou, M.; Bailey, D.; Barclay, W.S. The SARS-CoV-2 variant, Omicron, shows rapid replication in human primary nasal epithelial cultures and efficiently uses the endosomal route of entry. *bioRxiv* **2022**, 2021.2012.2031.474653, doi:10.1101/2021.12.31.474653.
63. Lyngse, F.P.; Kirkeby, C.T.; Denwood, M.; Christiansen, L.E.; Mølbak, K.; Møller, C.H.; Skov, R.L.; Krause, T.G.; Rasmussen, M.; Sieber, R.N.; et al. Household transmission of SARS-CoV-2 Omicron variant of concern subvariants BA.1 and BA.2 in Denmark. *Nat Commun* **2022**, *13*, 5760, doi:10.1038/s41467-022-33498-0.
64. Brown, J.C.; Goldhill, D.H.; Zhou, J.; Peacock, T.P.; Frise, R.; Goonawardane, N.; Baillon, L.; Kugathasan, R.; Pinto, A.L.; McKay, P.F.; et al. Increased transmission of SARS-CoV-2 lineage B.1.1.7 (VOC 2020212/01) is not accounted for by a replicative advantage in primary airway cells or antibody escape. *bioRxiv* **2021**, 2021.2002.2024.432576, doi:10.1101/2021.02.24.432576.
65. Mlcochova, P.; Kemp, S.A.; Dhar, M.S.; Papa, G.; Meng, B.; Ferreira, I.; Datir, R.; Collier, D.A.; Albecka, A.; Singh, S.; et al. SARS-CoV-2 B.1.617.2 Delta variant replication and immune evasion. *Nature* **2021**, *599*, 114-119, doi:10.1038/s41586-021-03944-y.
66. Saito, A.; Irie, T.; Suzuki, R.; Maemura, T.; Nasser, H.; Uriu, K.; Kosugi, Y.; Shirakawa, K.; Sadamasu, K.; Kimura, I.; et al. Enhanced fusogenicity and pathogenicity of SARS-CoV-2 Delta P681R mutation. *Nature* **2022**, *602*, 300-306, doi:10.1038/s41586-021-04266-9.
67. Benvenuto, D.; Angeletti, S.; Giovanetti, M.; Bianchi, M.; Pascarella, S.; Cauda, R.; Ciccozzi, M.; Cassone, A. Evolutionary analysis of SARS-CoV-2: how mutation of Non-Structural Protein 6 (NSP6) could affect viral autophagy. *J Infect* **2020**, *81*, e24-e27, doi:10.1016/j.jinf.2020.03.058.

68. Salpini, R.; Alkhatib, M.; Costa, G.; Piermatteo, L.; Ambrosio, F.A.; Di Maio, V.C.; Scutari, R.; Duca, L.; Berno, G.; Fabeni, L.; et al. Key genetic elements, single and in clusters, underlying geographically dependent SARS-CoV-2 genetic adaptation and their impact on binding affinity for drugs and immune control. *J Antimicrob Chemother* **2021**, *76*, 396-412, doi:10.1093/jac/dkaa444.
69. Wu, H.; Xing, N.; Meng, K.; Fu, B.; Xue, W.; Dong, P.; Tang, W.; Xiao, Y.; Liu, G.; Luo, H.; et al. Nucleocapsid mutations R203K/G204R increase the infectivity, fitness, and virulence of SARS-CoV-2. *Cell Host Microbe* **2021**, *29*, 1788-1801.e1786, doi:10.1016/j.chom.2021.11.005.
70. Müller, N.F.; Stolz, U.; Dudas, G.; Stadler, T.; Vaughan, T.G. Bayesian inference of reassortment networks reveals fitness benefits of reassortment in human influenza viruses. *Proc Natl Acad Sci U S A* **2020**, *117*, 17104-17111, doi:10.1073/pnas.1918304117.
71. Posada, D.; Crandall, K.A. The effect of recombination on the accuracy of phylogeny estimation. *J Mol Evol* **2002**, *54*, 396-402, doi:10.1007/s00239-001-0034-9.
72. Li, X.; Wang, W.; Zhao, X.; Zai, J.; Zhao, Q.; Li, Y.; Chaillon, A. Transmission dynamics and evolutionary history of 2019-nCoV. *J Med Virol* **2020**, *92*, 501-511, doi:10.1002/jmv.25701.
73. Li, X.; Zai, J.; Zhao, Q.; Nie, Q.; Li, Y.; Foley, B.T.; Chaillon, A. Evolutionary history, potential intermediate animal host, and cross-species analyses of SARS-CoV-2. *J Med Virol* **2020**, *92*, 602-611, doi:10.1002/jmv.25731.
74. Nie, Q.; Li, X.; Chen, W.; Liu, D.; Chen, Y.; Li, H.; Li, D.; Tian, M.; Tan, W.; Zai, J. Phylogenetic and phylodynamic analyses of SARS-CoV-2. *Virus Res* **2020**, *287*, 198098, doi:10.1016/j.virusres.2020.198098.
75. Li, Y.T.; Polotan, F.G.M.; Sotelo, G.I.S.; Alpino, A.P.A.; Dolor, A.Y.M.; Tujan, M.A.A.; Gomez, M.R.R.; Onza, O.J.T.; Chang, A.K.T.; Bautista, C.T.; et al. Lineage BA.2 dominated the Omicron SARS-CoV-2 epidemic wave in the Philippines. *Virus Evol* **2022**, *8*, veac078, doi:10.1093/ve/veac078.
76. Mallapaty, S. Where did Omicron come from? Three key theories. *Nature* **2022**, *602*, 26-28, doi:10.1038/d41586-022-00215-2.
77. Smyth, D.S.; Trujillo, M.; Gregory, D.A.; Cheung, K.; Gao, A.; Graham, M.; Guan, Y.; Guldenpfennig, C.; Hoxie, I.; Kannoly, S.; et al. Tracking cryptic SARS-CoV-2 lineages detected in NYC wastewater. *Nat Commun* **2022**, *13*, 635, doi:10.1038/s41467-022-28246-3.
78. Wei, C.; Shan, K.J.; Wang, W.; Zhang, S.; Huan, Q.; Qian, W. Evidence for a mouse origin of the SARS-CoV-2 Omicron variant. *J Genet Genomics* **2021**, *48*, 1111-1121, doi:10.1016/j.jgg.2021.12.003.
79. Kupferschmidt, K. Where did 'weird' Omicron come from? *Science* **2021**, *374*, 1179, doi:10.1126/science.acx9738.
80. Yamasoba, D.; Kimura, I.; Nasser, H.; Morioka, Y.; Nao, N.; Ito, J.; Uriu, K.; Tsuda, M.; Zahradnik, J.; Shirakawa, K.; et al. Virological characteristics of the SARS-CoV-2 Omicron BA.2 spike. *Cell* **2022**, *185*, 2103-2115.e2119, doi:10.1016/j.cell.2022.04.035.
81. Kumar, N.; Kaushik, R.; Tennakoon, C.; Uversky, V.N.; Longhi, S.; Zhang, K.Y.J.; Bhatia, S. Insights into the evolutionary forces that shape the codon usage in the viral genome segments encoding intrinsically disordered protein regions. *Brief Bioinform* **2021**, *22*, doi:10.1093/bib/bbab145.
82. Kumar, N.; Uversky, V.N.; Tomar, S.; Li, K.S.M.; Chappell, K.; Lau, S.K.P. Editorial: Emerging and re-emerging viral zoonoses. *Front Microbiol* **2022**, *13*, 978259, doi:10.3389/fmicb.2022.978259.