



Precise. Reliable.
Sleek and
Squeaky Clean.








Ensure excellence with PHCbi brand products.



PHCbi

Life Science Innovator Since 1966

Genome-based comparison between the recombinant SARS-CoV-2 XBB and its parental lineages

Fabio Scarpa¹  | Daria Sanna¹ | Ilenia Azzena^{1,2} | Marco Casu² |
Piero Cossu² | Pier Luigi Fiori^{1,3}  | Domenico Benvenuto⁴  | Elena Imperia^{4,5} |
Marta Giovanetti^{6,7}  | Giancarlo Ceccarelli⁸  | Roberto Cauda⁹ |
Antonio Cassone¹⁰ | Stefano Pascarella¹¹  | Massimo Ciccozzi⁴ 

¹Department of Biomedical Sciences, University of Sassari, Sassari, Italy

²Department of Veterinary Medicine, University of Sassari, Sassari, Italy

³Azienda Ospedaliera Universitaria (AOU) Sassari, Sassari, Italy

⁴Units of Medical Statistics and Molecular Epidemiology, University Campus Bio-Medico of Rome, Rome, Italy

⁵Unit of Gastroenterology, Department of Medicine, University Campus Bio-Medico of Rome, Rome, Italy

⁶Laboratório de Flavivírus, Instituto Oswaldo Cruz, Fundação Oswaldo Cruz, Rio de Janeiro, Brazil

⁷Department of Science and Technology for Humans and the Environment, University of Campus Bio-Medico di Roma, Rome, Italy

⁸Department of Public Health and Infectious Diseases, University Hospital Policlinico Umberto I, Sapienza University of Rome, Rome, Italy

⁹UOC Malattie Infettive, Department of Infectious Disease, Fondazione Policlinico Universitario Agostino Gemelli IRCCS, Rome, Italy

¹⁰Center of Genomic, Genetic and Biology, Siena, Italy

¹¹Department of Biochemical Sciences "A. Rossi Fanelli", Sapienza Università di Roma, Rome, Italy

Correspondence

Fabio Scarpa, Department of Biomedical Sciences, University of Sassari, Sassari, Italy.
Email: fscarpa@uniss.it

Funding information

PON "Ricerca e Innovazione" 2014–2020 (to Marta Giovanetti); CRP–ICGEB RESEARCH GRANT 2020 Project CRP/BRA20-03 (to Marta Giovanetti); FONDAZIONE DI SARDEGNA bando 2022–2023 for the Dipartimento di Scienze Biomediche - UNISS (to Daria Sanna and Fabio Scarpa); Sapienza grant no. RP12117A7670A1E8 (to Stefano Pascarella)

Abstract

Recombination is the main contributor to RNA virus evolution, and SARS-CoV-2 during the pandemic produced several recombinants. The most recent SARS-CoV-2 recombinant is the lineage labeled XBB, also known as Gryphon, which arose from BJ.1 and BM.1.1.1. Here we performed a genome-based survey aimed to compare the new recombinant with its parental lineages that never became dominant. Genetic analyses indicated that the recombinant XBB and its first descendant XBB.1 show an evolutionary condition typical of an evolutionary blind background with no further epidemiologically relevant descendant. Genetic variability and expansion capabilities are slightly higher than parental lineages. Bayesian Skyline Plot indicates that XBB reached its plateau around October 6, 2022 and after an initial rapid growth the viral population size did not further expand, and around November 10, 2022 its levels of genetic variability decreased. Simultaneously with the reduction of the XBB population size, an increase of the genetic variability of its first sub-lineage XBB.1 occurred, that in turn reached the plateau around November 9, 2022 showing

Fabio Scarpa and Daria Sanna contributed equally to this work.

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2023 The Authors. *Journal of Medical Virology* published by Wiley Periodicals LLC.

a kind of vicariance with its direct progenitors. Structure analysis indicates that the affinity for ACE2 surface in XBB/XBB.1 RBDs is weaker than for BA.2 RBD. In conclusion, at present XBB and XBB.1 do not show evidence about a particular danger or high expansion capability. Genome-based monitoring must continue uninterrupted to individuate if further mutations can make XBB more dangerous or generate new subvariants with different expansion capability.

KEYWORDS

coronavirus, epidemiology, pandemics, SARS coronavirus, virus classification

1 | INTRODUCTION

Currently, the world is being confronted with the ongoing pandemic of COVID-19 caused by the SARS-CoV2 identified for the first time in December 2019 during a pneumonia outbreak in Wuhan (China).¹ After the first cases, it rapidly emerged as a worldwide concern.² On March 11, 2020, with a total of 149 295 confirmed cases, the World Health Organization (WHO) declared COVID-19 a pandemic (<https://www.who.int/director-general/speeches/detail/who-director-general-s-opening-remarks-at-the-media-briefing-on-covid-19-11-march-2020>). In December 2022, WHO declared that, over 645 million confirmed cases and over 6.6 million deaths have been reported globally (<https://www.who.int/publications/m/item/weekly-epidemiological-update-on-covid-19-14-december-2022>).

Due to the continuing evolution of SARS-CoV-2 and the expected generation of new variants, SARS-CoV-2 infections are likely to remain a problem for the time being in most countries.³ Indeed, SARS-CoV-2 is a positive-sense single-stranded RNA virus, with a high error rate in RNA replication thus mutation and evolutionary fitness mostly affecting its transmissibility.⁴

Accordingly, during the pandemic SARS-CoV-2 significantly mutated over the time producing many lineages and sub-lineages with different expansion capabilities.⁵ For the generation of new viral variants, recombination events may play a nonnegligible role. Indeed, recombination is generally the main contributor to RNA virus evolution,⁶ as well as the re-assortment which, however, is present only in RNA viruses with segmented genomes such as the influenza virus (see e.g., Mugosa et al.⁷). Of course, recombination within different lineages requires the co-circulation and the co-infection of the viruses in the same host.⁸ As for the newly discovered variants, the occurrence recombinants must be monitored and requires a constant surveillance. The most recent SARS-CoV-2 recombinant is the lineage labeled as XBB, which has been also nicknamed Gryphon.⁹ XBB lineage is a recombinant of BJ.1 (also known as *Argus*) and BM.1.1.1 (also known as *Mimas*), both belonging to the BA.2 lineage (<https://www.who.int/activities/tracking-SARS-CoV-2-variants>), with the following Spike mutations in addition to those typical of its progenitor lineage: V83A, Y144-, H146Q, Q183E, V213E, G252V, G339H, R346T, L368I, V445P, G446S, N460K, F486S, F490S, and T11A (E), K47R (ORF1a), G662S (ORF1b), S959P

(ORF1b), and G8 (ORF8) as additional ones outside of the spike protein (<https://outbreak.info/compare-lineages>).¹⁰

As of December 2, 2022, from the sequences submitted to GISAID, XBB and its descendant showed a global sequence prevalence of around 7% (<https://gisaid.org/phylogenetics/global/nextstrain/>). The Technical Advisory Group on SARS-CoV-2 Virus Evolution discussed on the growth advantage of this sublineage and some early evidence on clinical severity and reinfection risk in several countries (i.a. Singapore and India) (<https://www.who.int/news/item/27-10-2022-tag-ve-statement-on-omicron-sublineages-bq.1-and-xbb>).

Recently, an alarming evidence has been provided on the capacity by XBB and XBB.1 of evading antibody-mediated immunity conferred by vaccination or previous infection due to the mutations or loss of B cell epitopes.¹¹ Notably, however, fully vaccinated subjects remain still protected from hospitalization and death owing to the preserved antiviral activity of T cells directed against conserved T epitopes present on the Spike protein.¹² Overall, XBB requires a constant genome-based epidemiological surveillance, immunological and clinical monitoring aimed at identification and functional evaluation of any mutation in the genome sequence potentially impacting virus transmissibility, host antibody and cell-mediated immune responses, and pathogenicity.

In such a context, here we performed a genome-based survey focused on genetic variability/phylogenetics and structural analyses to obtain an as complete as possible assessment of the evolutionary potential for epidemiological trajectory and dangers that XBB and its descendant XBB.1 could inflict to the population. In particular, the recombinant XBB and its first descendant XBB.1 have been compared with its parental lineages, to understand if these two lineages present new biological features that may candidate them to spread quickly and possibly outcompete the parental lineages.

2 | MATERIALS AND METHODS

To locate XBB and XBB.1 into an evolutionary perspective, genomic epidemiology of SARS-CoV-2 Omicron variants has been reconstructed by using a subsampling focused globally over the past 6 months, built with nextstrain/ncov (<https://github.com/nextstrain/>

ncov) (available at <https://gisaid.org/phyldynamics/global/nextstrain/>), including all genomes belonging to the GISAID Clade 21L (Omicron) (2371 of 2946 genomes sampled between January 2022 and December 2022).

After the first genomic assessment, to perform a genetic comparison between XBB and XBB.1 and their parental lineages, four subsets were built: BJ.1 ($n = 134$), BM.1.1.1 ($n = 70$) XBB ($n = 1602$), and XBB.1 ($n = 1841$) and for each dataset independently the above described genetic analyses were performed.

Genomes were aligned by using the algorithm L-INS-I implemented in Mafft 7.471,¹³ producing datasets of 29 729 (BJ.1), 29 719 (BM.1.1.1), 29 717 (XBB), and 29 725 (XBB.1) bp long. Manual editing was performed by using the software Unipro UGENE v.35.¹⁴ The software jModeltest 2.1.1¹⁵ was used to find the best probabilistic model of genome evolution with a maximum likelihood optimized search. Times of the most recent common ancestor and evolutionary rate were estimated by using Bayesian Inference (BI), which was carried out using the software BEAST 1.10.4¹⁶ with runs of 200 million generations under several demographic and clock model. To inference on the best representative output the selection of the better model was performed by the test of Bayes Factor¹⁷ by comparing the 2lnBF of the marginal likelihoods values following Mugosa et al.⁷ The software Beast was also used to draw the Bayesian Skyline Plot (BSP) and lineages thorough times for XBB, XBB.1 BJ.1, and BM.1.1.1 with runs of 200 million generations under the Bayesian Skyline Model with the uncorrelated log-normal relaxed clock model. For BSP analyses, for each analyzed lineage, all of the available genomes were used (see Table S1). Each included genome presented high quality, high coverage and full sampling date information. To avoid bias linked to a not fully randomized choice of genomes, several filters has been applied to ensure an equal distribution of the genetic variability between localities and sampling collection and the obtained dataset presented a balanced representation of the worldwide genetic variability for each analyzed lineage.

All datasets were built by downloading genomes form GISAID portal (<https://gisaid.org/>) available at November 16, 2022. See Table S1 for details on the used genomes and Authorship.

Break-point of the recombination event was individuated on a dataset sequence composed of all investigated genomes belong to the two parental lineages and the recombinant lineage (BJ.1 + BM.1.1.1 + XBB).

The mutations characterizing the XBB and XBB.1 SARS-CoV-2 Spike lineages were individuated by using consensus sequences obtained applying a cutoff of 75% sequence prevalence on all available sequences. The cutoff has been choice in accordance with the threshold used by GISAID lineage comparison tool (<https://gisaid.org/lineage-comparison/>).

After individuation, mutations were verified by comparing results of the “Lineage Comparison” web page of GISAID. Homology models of the mutant Spikes were created by means of the software Modeller 10.3.¹⁸ Spike structures were displayed and analyzed with the graphic program PyMOL.¹⁹ Surface electrostatic potential has been calculated with the program APBS²⁰ and displayed as

a two-dimensional projection with the SURFMAP software.²¹ SURFMAP implements a method of “molecular cartography” by means of which a protein three-dimensional surface can be projected onto a two-dimensional plane. Distribution of physico-chemical features over the protein surface can be analyzed and compared using the two-dimensional map. Net charge were calculated by the software PROPKA3.²² Foldx5²³ was applied to optimize the side chain conformation of the models built by Modeller using the Foldx5 function “RepairPDB.” Interaction energy between the Spike RBD and ACE2 were predicted with the Foldx5 function “AnalyseComplex.” Interface residue-residue interactions were assessed with the Foldx5 “PrintNetwork” function. In silico mutagenesis was obtained with the built-in functions available within PyMOL. In silico alanine scanning of the residues at the interface between RBD and ACE2 was carried out using the method available via the web server DrugScore (PPI).²⁴ The method is a fast and accurate computational approach to predict changes in the binding free energy when each residue at the subunit interface is mutated into alanine. The predicted net charge of the domains at pH 7.0 (set as a reference pH, though not necessarily reflecting the physiological environment) was calculated by means of the program PROPKA3. The predicted overall net charge is dependent on the combination of the side chain charges. The net charge is influenced by the interactions of the side chains with the surrounding atoms. To sample the variability of the interactions, 100 homology models of each RBD have been calculated with Modeller. Indeed, the Modeller refinement stage of the homology modeling can produce models differing for the conformation of side chains. Each model has been optimized by the Foldx5 “RepairPDB” procedure and the net charge has been calculated by PROPKA3. The final charge was the average of the 100 charges and the variability estimated by the standard error. To sample the fluctuations of the side chain conformations, the same procedure was applied to estimate the interaction energy of the complex ACE2-RBD.

3 | RESULTS

Phylogenomic reconstruction (Figure 1) indicates that XBB and XBB.1 (GISAID Clade 22F) genomes clustered within the not-monophyletic GISAID Clade 21L. More specifically, and as expected, they are evolutionary close to genomes of BA.2 which represent their ancient progenitor. See Table 1 for details on Nextstrain clade, Pango lineage, and WHO labels. Results of the Bayes Factor on the four datasets revealed that the Bayesian Skyline Model under the lognormal uncorrelated relaxed clock model fitted data significantly better than other tested demographic and clock models for all analyses datasets, with a value of 2lnBF = 32.9.

The Time of the Most Recent Common Ancestor of the clade composed of XBB + XBB.1 is placed 115 days before November 12, 2022, i.e., July 20, 2022, with a date interval confidence of 176–68 days (i.e., May 20 to September 5).

The breakpoint has been individuated between the nucleotidic position 22 901 and 22 939 in the SARS-CoV-2 reference genome

Phylogeny

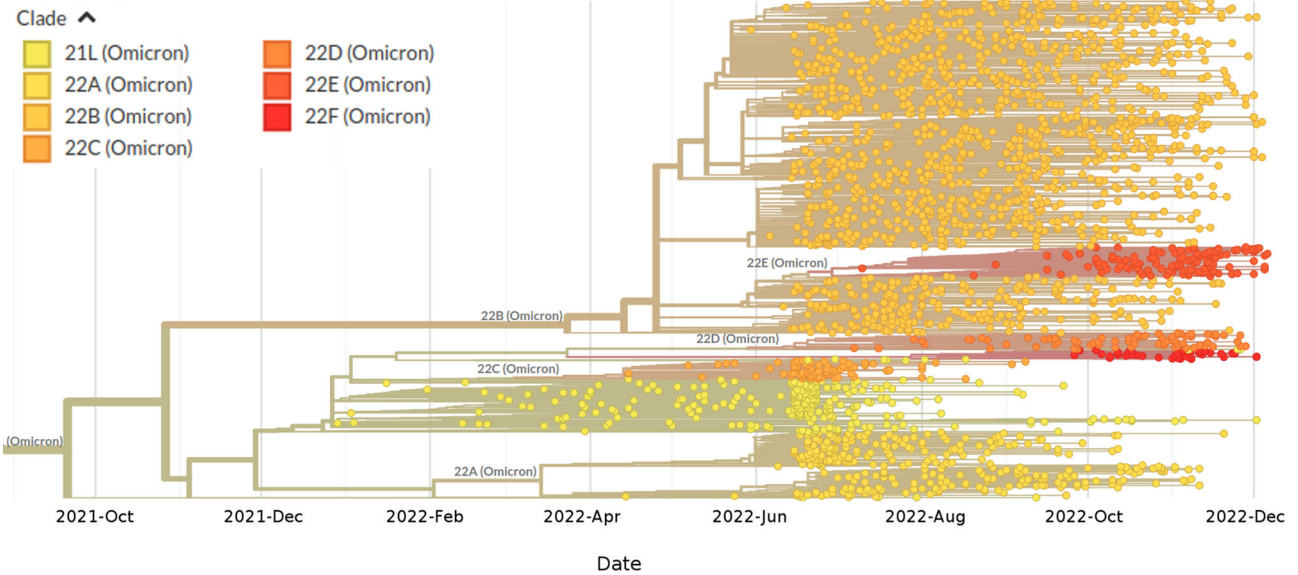


FIGURE 1 Highlight of the Omicron Clade (GISAID Clade 21L) in the time-scaled phylogenetic tree of a representative global subsample of 2371 SARS-CoV-2 genomes sampled between January 2022 and December 2022. Figures has been edited by using the software GIMP 2.8 (available at <https://www.gimp.org/downloads/oldstable/>). See Table 1 for details on Nextstrain clade, Pango lineage and WHO labels. WHO, World Health Organization.

TABLE 1 Nextstrain clade, Pango lineage, and WHO labels of the investigate lineages showed in Figure 1.

Nextstrain clade	Pango lineage	WHO label
21L (Omicron)	BA.2	o (Omicron)
22A (Omicron)	BA.4	o (Omicron)
22B (Omicron)	BA.5	o (Omicron)
22C (Omicron)	BA.2.12.1	o (Omicron)
22D (Omicron)	BA.2.75	o (Omicron)
22E (Omicron)	BQ.1	o (Omicron)
22F (Omicron)	XBB	o (Omicron)

Abbreviation: WHO, World Health Organization.

NC_045512.2. This region corresponds to the middle of the Receptor Binding Domain (see Figure 2).

BSP of the parental lineage BJ.1 (Figure 3A) showed that viral population has undergone an increase in size starting from about 110 days before October 28, 2022 (i.e., July 10, 2022), reaching the peak about 73 days before October 28, 2022 (i.e., August 16, 2022). Lineages through times plot (Figure 3B) indicates that the maximum number of lineages has been achieved around 73 days before October 28, 2022 (i.e., August 16, 2022).

BSP of the parental lineage BM.1.1.1 (Figure 3C) indicated that after the initial expansion in population size, the peak was reached around 91 days before November 3, 2022 (i.e., August 4, 2022) when started the plateau. Lineages through times plot (Figure 3D) indicates that the maximum number of lineages has been achieved around 100 days before November 3, 2022 (i.e., July 26, 2022).

BSP of the recombinant XBB (Figure 4A) showed that after a first period of flattened genetic variability, the viral population has undergone an increase in size starting from 46 days before November 12, 2022 (i.e., September 27, 2022), reaching the peak around 10 days after (i.e., October 6, 2022) and followed by a plateau phase that lasted until few days before November 12, 2022 when a reduction in viral population size occurred. Lineages through times plot (Figure 4B) indicates a moderately slow increase of the number of lineages until around 46 days before November 12, 2022 (i.e., September 27, 2022), then the number of lineages stopped its increasing.

BSP of the sublineage XBB.1 (Figure 4C) showed a brief period of flattened genetic variability after that an increase of the viral population size started around 70 days before November 9, 2022 (i.e., August 31, 2022) reaching its peak around 58 days before November 9, 2022 (i.e., September 12, 2022), then the plateau which is still ongoing without variation in genetic variability and viral population size. Lineages through times plot (Figure 4D) indicates a moderately slow increase of the number of lineages until around 64 days before November 9, 2022 (i.e., September 6, 2022), then the number of lineages stopped its increasing.

Evolutionary rate of the four tested lineages amount to 1.4×10^{-3} [95% HPD 6.4×10^{-4} – 2.4×10^{-3}], 1.3×10^{-3} [95% HPD 7.4×10^{-4} – 1.9×10^{-3}], 7.6×10^{-5} [95% HPD 4.57×10^{-5} – 1.0×10^{-4}] and 6.3×10^{-4} [95% HPD 5.2×10^{-4} – 7.3×10^{-4}] subs/sites/years for BJ.1, BM.1.1.1, XBB and XBB.1, respectively.

The structural analysis has been restricted to the most widespread lineages XBB and XBB.1 for which a wealth of sequential information is available in the databanks that allows a more precise

BJ.1 (BA.2.10.1) Donor	V 83	Y 144	H 146	*	*	*	Q 183	*	V 213	*	G 339	R 346	L 368	V 445	G 446	*	V 483	*	F 490
	A	-	Q				E		E		H	T	I	P	S		A		V
BM.1.1.1 (BA.2.75) Receptor	*	*	*	K 147	W 152	F 157	*	I 210	*	G 257	G 339	R 346	*	*	G 446	N 460	*	F 486	F 490
				E	R	L		V		S	H	T			S	K		S	S
XBB Clade 22F Recombinant	V 83	Y 144	H 146	*	*	*	Q 183	*	V 213	*	G 339	R 346	L 368	V 445	G 446	N 460	*	F 486	F 490
	A	-	Q				E		E		H	T	I	P	S	K		S	S

Breakpoint

FIGURE 2 Scheme showing the viral genome characterizing mutation of the recombinant XBB and parental lineages BJ.1 and BM.1.1. In the scheme it is possible to appreciate the breakpoint of the recombination event.

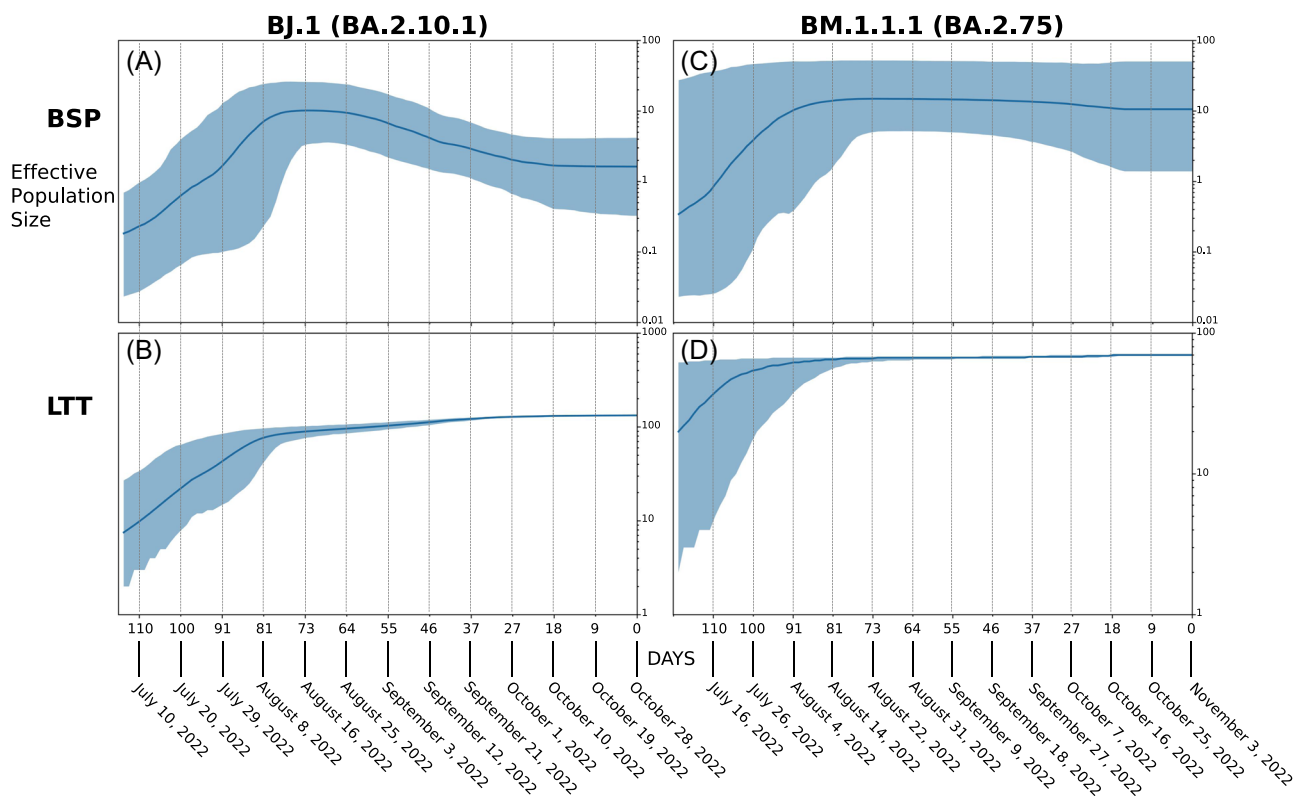


FIGURE 3 Bayesian skyline plot and lineages through time of SARS-CoV-2 BJ.1 (A, B) and BM.1.1.1 (C, D) variants. The viral effective population size (A, C) and the number of lineages (B, D) in the y-axis are shown as a function of days (x-axis).

definition of mutations prevalence. The characterizing XBB and XBB.1 mutations are reported in Table S2. The N-terminal domains (NTD) of the two variant Spikes differ for the mutation G252V that is present in XBB.1 and is missing in XBB. On the contrary, the two RDB domains share an identical mutational profile. The structural properties of the two variant Spikes have been compared to those of the parent BA.2 Spike.

The NTD mutations unique to XBB or XBB.1 with respect to BA.2 are displayed in Table S2. The mutation V83A occurs in a site within an exposed loop (Figure 5). The deletion of Y144 and the substitution H146Q are in the β -strand encompassed by the

sequence positions 140–146. The two mutation positions are near the putative binding site of the AXL receptor shown in Figure 5.^{25,26} The mutation G152V is specific of XBB.1 (Table S2) and it occurs within a loop also putatively involved in the interaction with the AXL receptor. The mutation Q183E occurs in an exposed loop and introduces a negatively charged residue that decreases the overall net charge and alters the distribution of the surface electrostatic potential (Table 2 and Figure 6A) with respect to the parental BA.2.

The XBB and XBB.1 RBD domains share the same mutation pattern (see Figure S1). The mutations that distinguish the two domains from BA.2 RBD are listed in Table S2. The mutation R346T

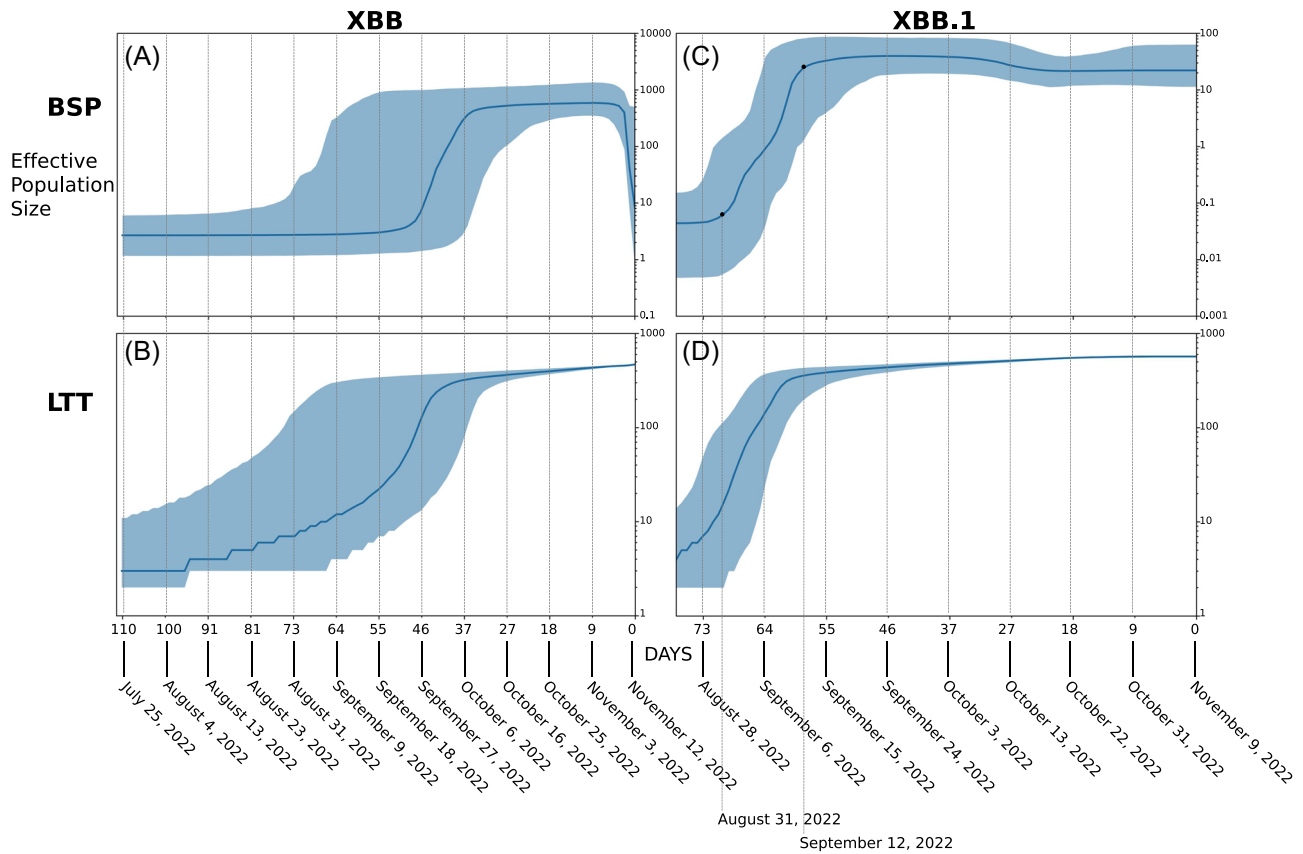


FIGURE 4 Bayesian skyline plot and lineages through time of SARS-CoV-2 XBB (A, B) and XBB.1 (C, D) recombinant. The viral effective population size (A, C) and the number of lineages (B, D) in the y-axis are shown as a function of days (x-axis).

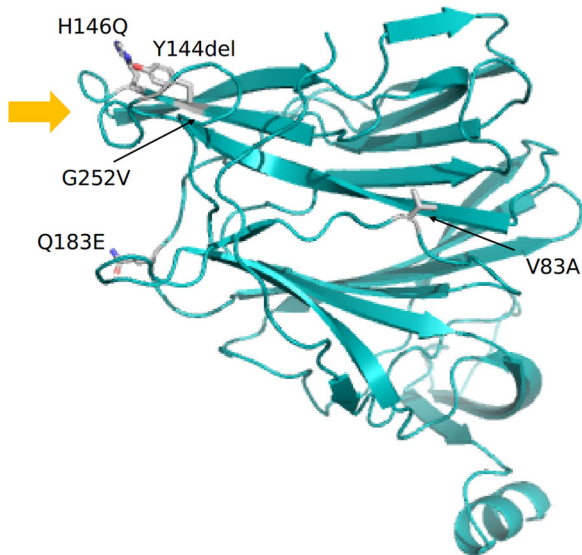


FIGURE 5 The N-terminal domain represented as a cartoon model. The mutated sites specific of XBB or XBB.1 are indicated by gray side chains reported as labeled stick models. Orange arrow marks the domain portion putatively involved in interaction with AXL.

TABLE 2 Net charge of RBD and NTD domains and interactions energy in the complex ACE2-RBD.

Variants	RBD ^a net charge	NTD net charge	RBD-ACE2 ^a (Kcal/mol)
BA.2	5.19 ± 0.01	0.95 ± 0.04	-6.19 ± 0.32
XBB	5.45 ± 0.02	-1.24 ± 0.04	-3.54 ± 0.30
XBB.1	=	-1.18 ± 0.03	=

Abbreviation: NTD, N-terminal domain.

^a"=" Means identical value.

occurs in a solvent exposed loop while L368I is within a short helix between positions 365–371 (Figure 7). The mutation V445P is in a loop near the interface to ACE2 although the side chain is not in direct contact with any ACE2 residue. On the contrary, G446S and F486S are in the ACE2 interface. The mutations N460K and F490S are also not directly involved in the ACE2 interface. Of note, BA.2 shows the mutation Q493S that is missing in XBB and XBB.1, in a site that is within the ACE2 interface.

The net charges of the BA.2 and XBB/XBB.1 are comparable although the positive surface potential distribution differs in the two

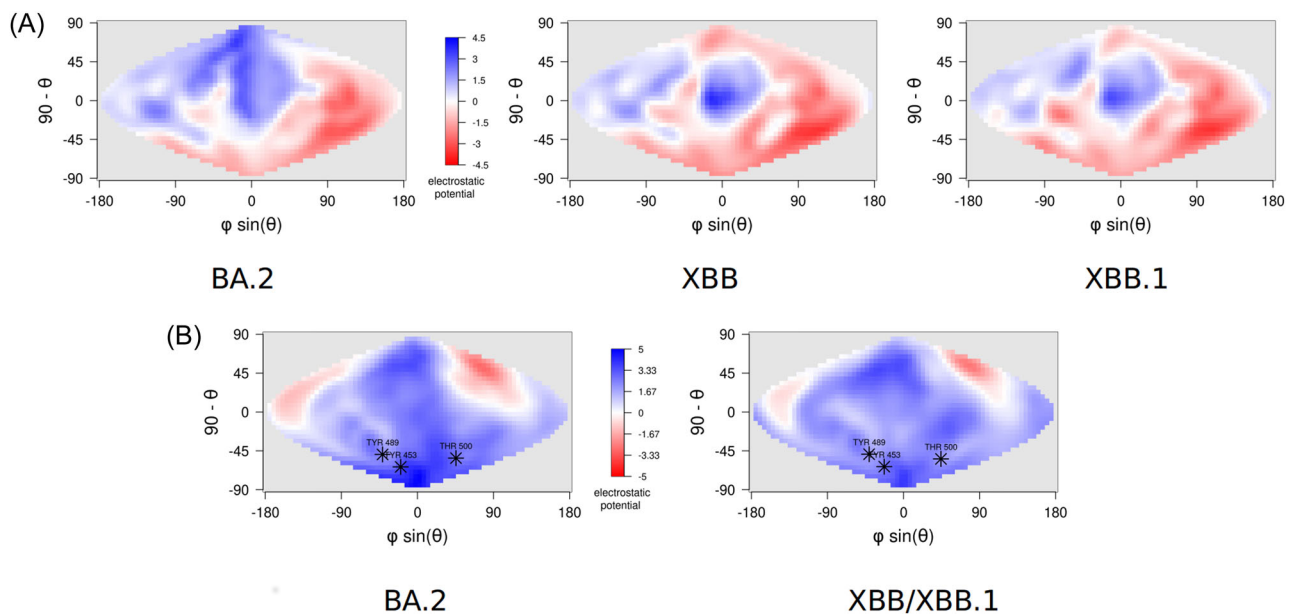


FIGURE 6 Projection on a two-dimensional map of the electrostatic potential surface of the NTDs of the three variants. Color scale is reported aside the BA.2 map. Electrostatic potential values are expressed as kT/e units. Map axes report the projected polar coordinates of the domains (A). Projection on a two-dimensional map of the electrostatic potential surface of the NTDs of the three variants. Color scale is reported aside the BA.2 map. Electrostatic potential values are expressed as kT/e units. Map axes report the projected polar coordinates of the domains. Asterisks map the position the three residues T453, Y489, and T500 belonging to the interface with ACE2 as a mark of the position of the interface on the projection (B). NTD, N-terminal domain.

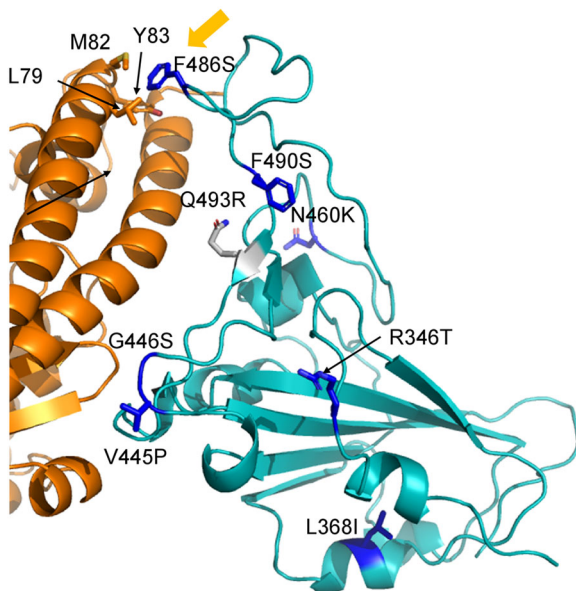


FIGURE 7 Portion of the complex between ACE2 (orange cartoon) and Spike RBD (teal cartoon). The XBB/XBB.1 specific mutated sites are highlighted with blue stick side chains labeled according to the corresponding mutation. The gray side chain marks the mutation specific to BA.2. ACE2 residues interacting with F486 are displayed as orange sticks and the corresponding region is highlighted by an orange arrow.

cases (Figure 6B). In fact, in the BA.2 RBD the positive potential seems to be more intense at the ACE2 interface area than in XBB/XBB.1. The prediction of the interaction energy between RBD and ACE2 suggests that BA2 RBD has a more stable interaction with this receptor than XBB/XBB.1 RBDs, within the accuracy limits of the predictive model. The putative weakening of the interactions may be in part explained by the XBB/XBB.1 mutations F486S. Indeed, alanine scanning carried out with the server DrugScore (PPI) predicts that F486 is an interface hot spot as its replacement with an Ala residue induces a loss of interaction energy of about 1.23 Kcal/mol. In the Wuhan RBD, F486 establishes hydrophobic interactions with the ACE2 residue L79, M82 and Y83.

4 | DISCUSSION

The SARS-CoV-2 XBB recombinant is the most recent product of a recombination event occurred during the current COVID-19 pandemic. As all newly discovered variants, XBB requires a stringest evaluation of its genomic differences from its parental lineages to estimate its capabilities for expansion and contagiousness as well as its pathogenicity features, including immunoevasion. Here we sought for a deep insight into the evolutionary and structural patterns of the SARS-CoV-2 recombinant XBB and its first descendant XBB.1 by

means a genome-based approach using of all genomes available in GISAID at November 16, 2022.

Phylogenomic reconstruction (Figure 1) indicates that genomes of XBB and XBB.1 (GISAID Clade 22F) belong to a monophyletic group, which in turn fall within the wide and heterogeneous GISAID Clade 22B as their parental lineages BM.1.1.1 and BJ.1. Although recombination odds are not linked to the evolutionary path, this is not surprising considering that they belong to the same sublineage (BA.2) and share a (not direct) common ancestor.

Similarly to what was recently observed for the variants BA.2.75 (nicknamed *Centaurus*)²⁷ and BQ.1 (nicknamed *Cerberus*),²⁸ the recombinant XBB and its first descendant XBB.1 show an evolutionary condition typical of an evolutionary blind background with no further epidemiologically relevant descendant that present features of concern.

In the near past the same condition was also observed in the BA.2.12.1 variant which indeed has not produced further new sub-lineages and whose genomic global sequence prevalence has been declining during time until its almost complete disappearance. Indeed, in the phylogenomic reconstruction, BA.2.12.1 variant (GISAID Clade 22C), BA.2.75 (GISAID Clade 22D), BQ.1 (GISAID Clade 22E) presented branches' length that suggested the lack of a rapid diversification,^{27,28} similarly to the clade composed of XBB + XBB.1 in the current reconstruction, which does not highlight any features typical of an epidemiologically dangerous lineage at the beginning of its evolutionary path.

The common ancestor to all analyzed genomes of XBB is temporally placed 115 days before November 12, 2022 (which is most recent collection date), i.e., July 20, 2022. This molecular dating calibration predates of about 1 month (23 days) the first detected genome of XBB (for which a complete sampling date is available), which was isolated in India on August 13, 2022. Phylogenomic reconstruction set a genome of BJ.1 from Bangladesh (EPI_ISL_14970755) external to the clade of XBB and XBB.1 in a basal position, while other member of BJ.1 are placed within another subcluster. The position in the phylogenomic tree of the genome of BJ.1 from Bangladesh, together with the high genome prevalence in Eastern India and Bangladesh of both XBB and its parental lineages, suggest the lack of multiple introductions and the occurrence of one recombination event (probably occurred in a human) during the long branch phase, before the parental lineages were in broad circulation. All these results, together with the genome prevalence of XBB occurred in early August (i.e., Bangladesh 92% and India 8%), suggest that the recombination event can be occurred in South Asia, where BJ.1 and BM.1.1.1 were very common. Indeed, in early August in Bangladesh the 76% of prevalence was represented by BJ.1 (43%) and BM.1.1.1 (33%), and in late August the prevalence of XBB was around to 35%. Within a month XBB (and its first descendant) became the first lineage in Bangladesh, by replacing its parental lineages with a prevalence of 93% in early October, and remaining almost the sole lineage until early December. It is interesting to note that this prevalence, partially reflect the worldwide growth of the lineage. Indeed, as represented by the BSP graph, the maximum viral

population size occurred between October 6, 2022 and early November 2022. This population dynamics further suggests that the origin of the recombinant lineage XBB is South Asia and allow to speculate that the most likely country of origin is Bangladesh.

During the recombination event the parental BJ.1 played the role of *Donor* while BM.1.1.1 is the *Acceptor* (*sensu* Focosi & Maggi⁸). The role of the parental lineages in the recombination event is further confirmed by the position of the breakpoint which is located in the genome between the nucleotidic position 22 901 and 22 939 in the SARS-CoV-2 reference genome (NC_045512.2), which correspond to the position 1339–1377 in the Spike protein gene sequence NC_045512.2 (21563–25384). More specifically the breakpoint occurred in the middle of the Receptor Binding Domain (around S:450–460), region for which the recombinant XBB carries the same mutation of BJ.1 until to G446S and from N640K carries the same mutations of BM.1.1.1.

As of today the recombinant lineage XBB appear to be common in Asia, but according to dating results here proposed, XBB circulated undisturbed for about a month before being detected and for several months before causing concern. As recently pointed out for several variants newly generated (i.a. BA.2.75 and BQ.1) this is not the feature of a highly expansive variant,^{27,28} which typically explodes much faster in terms of numbers of infections and population size, as it occurred for the Omicron variant (B.1.1.529) for instance, which became predominant in a few time.²⁹ Indeed, BSP indicates an initial period of about 60 days with a flattened genetic variability and a very small viral population size. In accordance with the increase of number of lineages depicted by the lineages through times plot, XBB showed an increase of its genetic variability and a consequent increase in population size started around September 27, 2022. After a rapid growing of the viral population size, the peak was reached in about 10 days and the plateau phase has begun around October 6, 2022. The period of quick growth has been an evolutionary advantage for XBB, which has been able to replace its parental lineages BJ.1 and BM.1.1.1 in Asia. However, it should be pointed out that as of today, worldwide, it has been never expanded and around November 10, 2022 its population size is reduced as a consequence of a lower level of genetic variability. In that period it happened a kind of vicariance with its first sub-lineage XBB.1, which indeed showed an increase of the number of lineage and of the viral population size shifted forward a few days in time. Indeed, XBB.1 reached its plateau around November 9, 2022 after a moderate increase of the number of lineage and genetic variability. These replacements, both that of XBB vs its parents and that of XBB.1 vs XBB, have been possible just because of the lack of a sufficient strength of the variant allowing it to prevail over all others. BJ.1 and BM.1.1.1 have never become dominant and their genome sequence prevalence was very low even before XBB arose. Notably, their evolutionary rate of 1.4×10^{-3} subs/sites/years for BJ.1 and 1.3×10^{-3} subs/sites/years for BM.1.1.1 are not so fast, as depicted by the BSP plots whose indicate that the peak phase has been reached around August 16, 2022 and July 26, 2022, respectively, with low level of increasing of lineages in both cases. This growth arrest of their population size allowed the increasing in

genomic sequence prevalence of XBB although it has a lower evolutionary rate (7.6×10^{-5} subs/sites/years). On the contrary XBB.1 easily replaced its direct progenitor thanks to a faster evolutionary rate of 6.3×10^{-4} subs/sites/years which present a difference equal to a factor of 10^{-1} .

The structural comparison between the three variants BA.2, XBB and XBB.1 suggests interesting differences. The NTDs of XBB and XBB.1 possess a more negative charge with respect to BA.2 that spreads over a large part of the molecular surface. Moreover, the characterizing mutations Y144del and H146Q occur near the putative site of interaction with the AXL receptor, that possesses an overall negative charge. In theory, these observations allow for the speculation that XBB and XBB.1 NTDs has a weaker propensity to interact with the AXL receptor and with the negatively charged sialosides displayed on the cell surface. Very likely, these changes also affect the interaction with the host immune system. Likewise, the mutations on XBB (identical in XBB.1) RBD do not significantly change the net charge that remains as much positive as in BA.2. However, the mutations alter the distribution of the positive electrostatic potential on the domain surface. In XBB/XBB.1 the positive potential becomes less localized at the ACE2 interface with respect to the BA.2 RBD. Moreover, the mutation F486S apparently destabilizes the complex ACE2-RBD. Once more, these observations suggest that XBB/XBB.1 RBDs may have a weaker affinity for ACE2 surface, such as it has been already described by Wang et al.³⁰ Moreover, the alterations of the physico-chemical properties of Spike induced by mutations may potentially modify the interaction with the host immune system components.

Most of the mutations of the SARS-CoV-2 variants analyzed in this article fall in the NTD, RBD and RBM domains while most of the S2 region is highly conserved. This could probably enforce the hypothesis that NTD, RBD and RBM regions of the spike protein contain numerous strong B cell epitopes with the ability to readily elicit a strong neutralizing antibody response. Accordingly, their numerous mutations are considered to be largely responsible of the capacity of XBB and XBB.1 variants of massively evading neutralization by antibodies in both the vaccinated and the infected subjects as well.¹¹ Moreover, it has been suggested that Spike glycoprotein could conceal each of its immunodominant domains by adopting the closed conformation.³¹

5 | CONCLUSIONS

In conclusion, the genome-based survey of the SARS-CoV-2 recombinant XBB and on its first sublineage XBB.1 suggest that, although this new lineage presents several spike mutations of interest, and overall highly immune-evasive capacity to escape from neutralizing antibodies,³² it currently does not show evidence of high expansion capabilities and/or a contagiousness rate higher than other ongoing circulating lineages. Recombination is a common evolutionary tool in the family Coronaviridae,³³ as well as for RNA viruses in general.⁸ As of now, XBB represents the first recombinant with

showed a growth advantage that deserved to be investigated. Indeed, by becoming regionally dominant, it has caused concern, suggesting the need for an in-depth survey. Nonetheless, its expansion capability appears to be very limited with the peak reached in October 6 for XBB and November 9 for XBB.1. Data here reported indicates an initial quick growth followed by a long period of flattened genetic variability, very distant, in term of expansion capabilities, from an epidemiologically dangerous lineage as shown at the beginning of the pandemic where population size presented an extremely vertical curve (see, i.a., Lai et al.³⁴).

The genome-based surveillance must continue for all SARS-CoV-2 lineages and variants to detect any new possible expansion. Structural interpretation of mutations provides a mean to formulate hypothesis for explaining and predicting the epidemiological behavior of the variants. New further mutations can make XBB more dangerous or generate new subvariants and accordingly, in the near future the attention must be focused on its descendant, to verify their expansion capability and biological features for a better understanding of the pandemic.

AUTHOR CONTRIBUTIONS

Conceptualization: Fabio Scarpa, Daria Sanna, Marco Casu, Pier Luigi Fiori, Massimo Ciccozzi. *Data Analyses:* Fabio Scarpa, Stefano Pascarella. *Writing-original draft preparation:* Fabio Scarpa, Daria Sanna, Marco Casu. *Writing-review and editing:* Fabio Scarpa, Daria Sanna, Ilenia Azzena, Marco Casu, Piero Cossu, Pier Luigi Fiori, Domenico Benvenuto, Elena Imperia, Marta Giovanetti, Giancarlo Ceccarelli, Roberto Cauda, Antonio Cassone, Stefano Pascarella, Massimo Ciccozzi. *Resources:* Fabio Scarpa, Daria Sanna, Marco Casu.

ACKNOWLEDGMENTS

This research was funded from by FONDAZIONE DI SARDEGNA bando 2022-2023 for the Dipartimento di Scienze Biomediche–UNISS (to Fabio Scarpa and Daria Sanna). Marta Giovanetti is funded by PON “Ricerca e Innovazione” 2014–2020 and by the CRP–ICGEB RESEARCH GRANT 2020 Project CRP/BRA20-03, contract CRP/20/03, Oswaldo Cruz Foundation. Stefano Pascarella is in part supported by the Sapienza grant no. RP12117A7670A1E8. The authors are grateful to the family of Prof. Giuseppe De Feo for the support. We also would like to thank all the authors who have kindly deposited and shared genomes on GISAID. Open Access Funding provided by Università degli Studi di Sassari within the CRUI-CARE Agreement.

CONFLICTS OF INTEREST STATEMENT

The authors declare no conflicts of interest.

DATA AVAILABILITY STATEMENT

Genomes analysed in the present study were taken from GISAID database and are available at <https://gisaid.org/>.

ORCID

Fabio Scarpa  <http://orcid.org/0000-0002-3501-714X>

Pier Luigi Fiori  <http://orcid.org/0000-0001-6190-612X>

Domenico Benvenuto  <http://orcid.org/0000-0003-3833-2927>
 Marta Giovanetti  <http://orcid.org/0000-0002-5849-7326>
 Giancarlo Ceccarelli  <http://orcid.org/0000-0001-5921-3180>
 Stefano Pascarella  <http://orcid.org/0000-0002-6822-4022>
 Massimo Ciccozzi  <http://orcid.org/0000-0003-3866-9239>

REFERENCES

- Zhou P, Yang XL, Wang XG, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature*. 2020;579(7798):270-273. doi:10.1038/s41586-020-2012-7
- Wu D, Wu T, Liu Q, Yang Z. The SARS-CoV-2 outbreak: what we know. *Int J Infect Dis*. 2020;94:44-48. doi:10.1016/j.ijid.2020.03.004
- Zella D, Giovanetti M, Benedetti F, et al. The variants question: what is the problem. *J Med Virol*. 2021;93:6479-6485. doi:10.1002/jmv.27196
- Hillen HS, Kovic G, Farnung L, Dienemann C, Tegunov D, Cramer P. Structure of replicating SARS-CoV-2 polymerase. *Nature*. 2020;584:154-156. doi:10.1038/s41586-020-2368-8
- Borsetti A, Scarpa F, Maruotti A, et al. The unresolved question on COVID-19 virus origin: the three cards game? *J Med Virol*. 2022;94:1257-1260. doi:10.1002/jmv.27519
- Lai MM. Genetic recombination in RNA viruses. *Curr Top Microbiol Immunol*. 1992;176:21-32. doi:10.1007/978-3-642-77011-1_2
- Mugosa B, Vujosevic D, Ciccozzi M, et al. Genetic diversity of the haemagglutinin (HA) of human influenza A (H1N1) virus in Montenegro: focus on its origin and evolution: genetic diversity of H1N1 in Montenegro. *J Med Virol*. 2016;88:1905-1913. doi:10.1002/jmv.24552
- Focosi D, Maggi F. Recombination in coronaviruses, with a focus on SARS-CoV-2. *Viruses*. 2022;14:1239. doi:10.3390/v14061239
- Focosi D, McConnell S, Casadevall A. The Omicron variant of concern: diversification and convergent evolution in spike protein, and escape from anti-spike monoclonal antibodies. *Drug Resist Updates*. 2022;65:100882. doi:10.1016/j.drug.2022.100882
- Gangavarapu K, Latif AA, Mullen J, et al. outbreak.info (available at <https://outbreak.info/compare-lineages>).
- Imai M, Ito M, Kiso M, et al. Efficacy of antiviral agents against Omicron Subvariants BQ.1.1 and XBB. *N Engl J Med*. 2022;388:89-91. <https://www.nejm.org/doi/10.1056/NEJMc2214302>
- Meyer S, Blaas I, Bollineni RC, et al. Prevalent and immunodominant CD8 T cell epitopes are conserved in SARS-CoV-2 variants. *Cell Rep*. 2023;42(1):111995. doi:10.1016/j.celrep.2023.111995
- Katoh K, Standley DM. MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Mol Biol Evol*. 2013;30:772-780. doi:10.1093/molbev/mst010
- Okonechnikov K, Golosova O, Fursov M, UGENE Team. Unipro UGENE: a unified bioinformatics toolkit. *Bioinformatics*. 2012;28:1166-1167. doi:10.1093/bioinformatics/bts091
- Darriba D, Taboada GL, Doallo R, Posada D. jModelTest 2: more models, new heuristics and parallel computing. *Nature Methods*. 2012;9:772. doi:10.1038/nmeth.2109
- Drummond AJ, Rambaut A. BEAST: Bayesian evolutionary analysis by sampling trees. *BMC Evol Biol*. 2007;7:214. doi:10.1186/1471-2148-7-214
- Kass RE, Raftery AE. Bayes factors. *J Am Stat Assoc*. 1995;90:773-795. doi:10.2307/2291091
- Webb B, Sali A. Protein structure modeling with MODELLER. *Methods Mol Biol*. 2017;1654:39-54. doi:10.1007/978-1-4939-0366-5_1
- Schrodinger LLC The PyMOL Molecular Graphics System, Version 1.8.9-54. 2015.
- Jurrus E, Engel D, Star K, et al. Improvements to the APBS biomolecular solvation software suite. *Prot Sci*. 2018;27(1):112-128. doi:10.1002/pro.3280
- Schweke H, Mucchielli MH, Chevrollier N, Gosset S, Lopes A. SURFMAP: a software for mapping in two dimensions protein surface features. *J Chem Inf Model*. 2022;62(7):1595-1601. doi:10.1021/acs.jcim.1c01269
- Olsson MHM, Søndergaard CR, Rostkowski M, Jensen JH. PROPKA3: consistent treatment of internal and surface residues in empirical pKa predictions. *J Chem Theory Comput*. 2011;7:525-537. doi:10.1021/ct100578z
- Delgado J, Radusky LG, Cianferoni D, Serrano L. FoldX 5.0: working with RNA, small molecules and a new graphical interface. *Bioinformatics*. 2019;35:4168-4169. doi:10.1093/bioinformatics/btz184
- Krieger DM, Gohlke H. DrugScorePPI webserver: fast and accurate in silico alanine scanning for scoring protein-protein interactions. *Nucleic Acids Res*. 2010;38:W480-W486. doi:10.1093/nar/gkq471
- Zeng C, Ye Z, Fu L, Ye Y. Prediction analysis of porcine AXL protein as a potential receptor for SARS-CoV-2. *J Infect*. 2022;84:579-613. doi:10.1016/j.jinf.2021.12.038
- Wang S, Qiu Z, Hou Y, et al. AXL is a candidate receptor for SARS-CoV-2 that promotes infection of pulmonary and bronchial epithelial cells. *Cell Res*. 2021;31(2):126-140. doi:10.1038/s41422-020-00460-y
- Scarpa F, Sanna D, Azzena I, et al. On the SARS-CoV-2 BA.2.75 variant: a genetic and structural point of view. *J Med Virol*. 2023;95:28119. doi:10.1002/jmv.28119
- Scarpa F, Sanna D, Benvenuto D, et al. Genetic and structural data on the SARS-CoV-2 Omicron BQ.1 variant reveal its low potential for epidemiological expansion. *Int J Mol Sci*. 2022;23:15264. doi:10.3390/ijms232315264
- Liu L, Iketani S, Guo Y, 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
- Wang Q, Iketani S, Li Z, et al. Alarming antibody evasion properties of rising SARS-CoV-2 BQ and XBB subvariants. *Cell*. 2023;186(2):279-286.e8. doi:10.1016/j.cell.2022.12.018
- Duan L, Zheng Q, Zhang H, Niu Y, Lou Y, Wang H. The SARS-CoV-2 spike glycoprotein biosynthesis, structure, function, and antigenicity: implications for the design of Spike-Based vaccine immunogens. *Front Immunol*. 2020;11:576622. doi:10.3389/fimmu.2020.576622
- Kurhade C, Zou J, Xia H, et al. Low neutralization of SARS-CoV-2 Omicron BA.2.75.2, BQ.1.1, and XBB.1 by parental mRNA vaccine or a BA.5-bivalent booster. *Nat Med*. 2023;29:344-347. doi:10.1038/s41591-022-02162-x
- Scarpa F, Sanna D, Azzena I, et al. Update on the phylodynamics of SARS-CoV. *Life*. 2021;11:820. doi:10.3390/life11080820
- Lai A, Bergna A, Acciarri C, Galli M, Zehender G. Early phylogenetic estimate of the effective reproduction number of SARS-CoV-2. *J Med Virol*. 2020;92:675-679. doi:10.1002/jmv.25723

SUPPORTING INFORMATION

Additional supporting information can be found online in the Supporting Information section at the end of this article.

How to cite this article: Scarpa F, Sanna D, Azzena I, et al. Genome-based comparison between the recombinant SARS-CoV-2 XBB and its parental lineages. *J Med Virol*. 2023;95:e28625. doi:10.1002/jmv.28625