Editor summary:

Evolution of Omicron lineages and future evolution trajectories of SARS-CoV-2.

Reviewer Recognition:

Nature Microbiology thanks Rowena Bull, Frank Kirchhoff and the other, anonymous, reviewer(s) for their contribution to the peer review of this work.

SARS-CoV-2 evolution in the Omicron era

Cornelius Roemer[1], Daniel J. Sheward[2], Ryan Hisner[3], Federico Gueli[4], Hitoshi Sakaguchi[5], Nicholas Frohberg[6], Josette Schoenmakers[7], Kenta Sato[2], Áine O'Toole[8], Andrew Rambaut[8], Oliver G. Pybus[9,10], Christopher Ruis[11,12,13], Ben Murrell[2], and Thomas P. Peacock[14, 15*]

[1]Biozentrum, University of Basel, Basel, Switzerland; Swiss Institute of Bioinformatics
[2]Department of Microbiology, Tumor and Cell Biology, Karolinska Institutet, Stockholm, Sweden
[3]University of Cape Town, Rondebosch, South Africa
[4]Independent Researcher, Italy
[5]Retired, Japan
[6]Independent Researcher, USA
[7]Independent Researcher, the Netherlands
[8]Institute of Ecology and Evolution, University of Edinburgh, UK
[9]Department of Biology, University of Oxford, UK
[10]Department of Pathobiology and Population Science, The Royal Veterinary College, London, UK
[11]Molecular Immunity Unit, University of Cambridge Department of Medicine, MRC-Laboratory of Molecular Biology, Cambridge, UK
[12]Department of Veterinary Medicine, University of Cambridge, Cambridge, UK
[13]Cambridge Centre for AI in Medicine, University of Cambridge, Cambridge, UK
[14]Department of Infectious Disease, Imperial College London, UK
[15]The Pirbright Institute, UK
*corresponding author: tpp09@ic.ac.uk

Since SARS-CoV-2 lineage BA.5 (Omicron) emerged and spread in 2022, Omicron lineages have massively diversified. Here, we review the evolutionary trajectories and processes underpinning the emergence of these lineages, and identify the most prevalent sublineages. We discuss the potential origins of second-generation BA.2 lineages. Simple and complex recombination, antigenic drift and convergent evolution have enabled SARS-CoV-2 to accumulate mutations that alter its antigenicity. We also discuss the likely future evolution trajectories of SARS-CoV-2.

Introduction

SARS-CoV-2 has been intensely sampled and sequenced, and is now a paradigm for understanding viral emergence and evolution in real time during a pandemic. Towards the end of 2020, and throughout 2021 and 2022, SARS-CoV-2 evolved rapidly and diversified into many lineages[1-3], with variants of concern (VOCs) identified using Greek letters such as Alpha or Omicron. Much of the lineage diversification has occurred in the viral spike protein, in part due to strong selection pressure from neutralising antibodies. Indeed, following *in vitro* selection with pooled sera from convalescent and vaccinated individuals, ancestral spikes evolved several antigenic mutations that were shared with the Omicron variant[4,5]. Throughout the pandemic SARS-CoV-2 variants have shown strong evidence of convergent evolution[6], where the same (or functionally similar) mutations have arisen multiple times, independently, in distinct genetic backgrounds. Many of the observed mutations in the SARS-CoV-2 spike protein are also predictable with deep mutational scanning, based on the reduction in antibody binding they confer[5,7], or enhancement in ACE2 binding or RBD stability[8,9].

After the initial discovery of the SARS-CoV-2 Omicron VOC in late 2021[10], many countries and regions experienced successive waves of infections caused by Omicron lineages and sublineages[1], including BA.1, BA.2, and BA.5, together with their sublineages. BA.1-5 are thought to have emerged from a shared but currently uncharacterised human or animal reservoir of Omicron genetic diversity[10,11]. From early 2023 a mix of Omicron sub-lineages are emerging and growing, often coinciding with waves of infections (for how new variants are identified, see box 1). Here we review the evolutionary events and processes that have led to the emergence of these lineages.

Second-generation variants

SARS-CoV-2 has periodically produced ‘variants’ defined by long ancestral phylogenetic branches, that lack genetic intermediates, and sometimes have an ancestral branch (for the variant) that emerges from inside older, not contemporary, genetic variation[3,10-14]. Interestingly, this feature may have parallels with the evolution of pandemic norovirus strains[15]. This seemingly ‘saltatory’ pattern of virus molecular evolution is

hypothesised to be the result of virus evolution during long-term chronic infections[3,13,14,16-18], with the evolution appearing to be saltatory due to a lack of intermediate sequences being collected from the infected individual (a review of the evidence for this hypothesis can be found in Markov et al[3]). The extended duration of some chronic infections might enable intrahost accumulation and fixation of mutations, owing to the long period of time within a single host and therefore no transmission bottleneck. Other hypotheses have been proposed, such as re-emergence of lineages from animal reservoirs, or cryptic circulation of SARS-CoV-2 in undersampled or geographically isolated regions of the world[3,19]. Whilst these hypotheses are biologically plausible and cannot be fully rejected, we believe that existing evidence points to chronic infections as the primary cause of extremely different variant lineages. Examples of variants that are extremely different include the variants of concern Alpha, Beta, Gamma and the Omicron lineages (e.g. BA.1, BA.2 and BA.5)[10,12-14,20].

The first generation of SARS-CoV-2 variants of concern (Alpha, Beta, Gamma, Delta and Omicron) each evolved from pre-VOC ancestors and sequentially replaced one another, either locally or globally[1]. In contrast, the Omicron lineages that dominated transmission in many locations during late 2022 typically descend from a BA.2 background and therefore represent 'second-generation' variants that evolved from a pre-existing VOC (Figure 1). These 'second-generation' BA.2 lineages generally contain 10-30 additional private mutations compared to their closest known ancestor. Examples include BA.2.75[21], BA.2.10.4, BJ.1, BS.1[22], BA.2.3.20, BA.2.83, BP.1, DD.1, BA.2.3.22, and most recently BA.2.86[23]. Like the first-generation variants, these Omicron sublineages carry numerous non-synonymous mutations, particularly in the spike protein N-terminal domain (NTD) and receptor binding domain (RBD) (Figure 1)[2].

No clear genetic intermediates between these second-generation variants and their Omicron precursors have been sampled, suggesting they evolved through a saltatory-like evolutionary process, such as in the setting of chronic infections, seeded near the end of 2021 or start of 2022. Intriguingly many of these 2nd generation BA.2 variants were first found in countries that had large BA.2, rather than BA.1 waves in late 2021/early 2022, for example India, the Philippines and Vietnam[24].

Until the takeover of XBB sublineages in early 2023 , BA.2.75 was the most widespread of the second-generation BA.2 variant lineages (Figure 2B)[25]. BA.2.3.20 also showed some appreciable growth in late 2022, particularly in the Philippines where it was the dominant lineages sequenced over a 6 month period in late 2022 into early 2023. As of September 2023, BA.2.86, appears to be growing in several regions. BA.2.86 contains over 40 mutations relative to BA.2[23], a similar range to that seen in the original Omicron BA.1 lineage relative to its B.1.1 lineage ancestor[10].

'Simple' recombination

Recombination is common in coronaviruses. Ever since there has been enough genetic diversity sequenced to unambiguously identify chimeric genomes, it has been clear that co-infection with different SARS-CoV-2 variants and homologous recombination between coinfecting viruses is possible[26]. During earlier periods of rapid lineage replacement, recombinant SARS-CoV-2 lineages were mainly identified when a prior wave was in steep decline, and a genetically distinct variant was emerging[26,27]. Recombinant lineages were often outcompeted by their parental lineages, arising too late to make a large impact, and appearing to have too little growth advantage relative to their parental lineages. However, in common with recombination in other viruses, recombinants between divergent SARS-CoV-2 variants can acquire unique advantageous properties from both parents[28]. As of late February 2023, there were 63 Pango-designated recombinant lineages, denoted by their 'X' prefix[29]. Recombinant Pango lineages are only designated with an X if they exhibit substantial spread, so this is likely a vast underestimate of the total number of recombinant lineages.

XBB is the most widespread inter-lineage recombinant to date, and is presently (as of September 2023) dominant worldwide[30] (Figure 2B). XBB is a recombinant between two second-generation BA.2 lineages: BJ.1 and a sublineage of BA.2.75 (most likely BM.1.1.1) (Figure 3). XBB inherited the 5' part of its genome from BJ.1 and the 3' end of its genome from BA.2.75, with a single breakpoint within the RBD of Spike. This Spike breakpoint resulted in the inheritance of advantageous antibody escape mutations from both BJ.1 (R346T, V445P, G446S) and BM.1.1.1 (N460K, F486S, F490S and the R493Q reversion), creating a highly distinctive combinations of antigenic RBD mutations and forming a spike highly resistant to previously generated neutralising antibodies[7]. Another notable simple recombinant lineages

is XBF[31] - a recombinants between BA.5 and BA.2.75 sublineages (Figure 3). More recently the first 'second order' recombinant lineage between the recombinant XBB.1 and a BA.2.75 sublineage was identified and designated XBL [32] (Figure 3).

'Complex' recombination

In addition to 'simple' recombinants, more 'complex' recombinants emerged and were detected for the first time in 2022[1,33,34]. Earlier SARS-CoV-2 recombinants were the result of recombinant events between extant, co-circulating lineages, and tended to contain 1 or 2 detectable breakpoints[26]. Complex recombinants have been generated from parental lineages not known to have co-circulated widely (e.g. Delta and BA.2 in the complex recombinants shown in Figure 3.), and typically contain greater numbers of breakpoints (between three and eight in these examples). They also carry many more 'private mutations' (mutations that do not appear to have been inherited from either parental lineage) than 'simple' recombinants (Figure 3). Furthermore, the complex recombinants XAY and XBA share parts of their genomes and private mutations with one another, suggesting they arose from a common ancestor[33]. Other examples of complex recombinants are lineages XAW (which has only 2 breakpoints) and XBC[34].

Due to the often high number of breakpoints and private mutations carried by these complex recombinants, and the fact that they have a non-contemporary parental lineage (one that has not been observed for some time in the region of sampling), we hypothesise they may arise during long-term chronic infections. This is consistent with the shared mutations of XAY and XBA and their near-simultaneous emergence in the same region, suggesting that they might have arisen within the same individual[33], similar to how previous chronic infections have shown huge intrahost diversity[35]. In the case of XAY, XBA, XBC, and XAW, it is possible that the chronically-infected individual was first infected by Delta and subsequently superinfected with BA.2 at a later date.

Of these complex recombinants, XBC and XAY were the most widespread (at least by numbers of genomes submitted as of July 2023) and continue to be sampled. XBC.1.6, an XBC sublineage with several additional antigenic changes (R346S and L452R), continues to show competitive growth advantages relative to the globally dominant XBB sublineages (Figure 2A).

Antigenic Drift

By mid-2022, BA.5 had become the predominant variant globally, displacing BA.2 in most regions[25]. Unlike previous VOCs or BA.2 lineages, which showed relatively little accumulation of antigenic mutations once they predominated (with the closest examples being BA.2.12.1 or BA.1.1), BA.5 instead began to accumulate antigenic mutations in a stepwise manner. This stepwise evolution contrasts with the initial second-generation BA.2 lineages, which lacked sampled intermediate sequences. Notable examples are the sublineages of BQ.1, the most widespread of which (in terms of genomes) was BQ.1.1 and its progeny, which contain three further antigenic mutations in the spike receptor binding domain (RBD) – the main target for neutralising antibodies in SARS-CoV-2 (Figure 1). Several examples with fewer antigenic mutations that showed some growth at the time included BA.4.6, BF.7 and BQ.1.1's parental lineage, BQ.1[30].

Furthermore, some of the aforementioned second-generation variants derived from BA.2 also show antigenic evolution. In particular BA.2.75 had many descendent sublineages that have accumulated antigenic RBD mutations through a stepwise antigenic drift-like process (Figure 1). Notable examples include BA.2.75.2, BR.2, BN.1.2.1, BM.1.1.1 and CH.1.1, all of which contain several additional antigenic RBD mutations compared to the parental BA.2.75 lineage (Figure 1)[7,30]. This 'drift'-like evolutionary pattern is consistent with the evolutionary processes seen in many other respiratory viruses, such as the stepwise antigenic drift in some seasonal coronaviruses and influenza viruses[36,37].

Immune-escape mutations can come at a cost to replicative fitness, and as a result are often accompanied by compensatory mutations. In the case of SARS-CoV-2, the RBD is the dominant target of neutralizing antibodies, and escape within these epitopes also has consequences for the affinity to the host receptor, ACE2. In line with this, a hypothesis has been developed to help explain why some SARS-CoV-2 lineages appear to be more tolerant of antigenic drift than others. This hypothesis states that lineages with relatively stronger ACE2 affinity are better able to tolerate antigenic mutations that result in slight reductions in ACE2 binding[38,39]. For example, BA.2.75, which has very strong ACE2 binding[39], rapidly accumulated antigenic mutations upon its circulation (Figure 1). Other recent examples include XBB.1.5, which also shows strong ACE2 binding, and is beginning to show similar

197 diversification in the RBD[39], and BA.2.3.20, which quickly gained antigenic mutations - G446S
198 and F486S - as the sublineage CM.8.1 (Figure 1).

199 **Convergent evolution**

200 One feature common to all the lineages discussed in the Perspective is the high degree
201 of convergent evolution they exhibit[7,40]. While some of these sites may enhance ACE2 binding
202 affinity (notably N460K, F486P and R493Q)[8,39,41], the majority are known or predicted to be
203 key antibody escape mutations[7,24,42]. Examples of antigenic substitutions, such as R346X
204 (where X represents any other amino acid), K444X, G446X, L452X, N460K, F486X (particularly
205 F486P, a 2-nucleotide change), F490X and the R493Q reversion are present in many of these
206 lineages (Figure 1)[24,40]. In recent months further convergent evolution has occurred within
207 dominant XBB sublineages, most prominently at RBD positions K356T, R403K, L455F, F456L,
208 Y453F and T478R. Several NTD changes, particularly deletions in the ~144 region of the NTD
209 supersite - the major target of antibodies in the NTD[43] - also appeared in BA.5 sublineages,
210 for example BQ.1.1.20, BQ.1.8 and BQ.1.23[7]. Similar NTD deletions are also found in BJ.1 (and
211 therefore XBB), BS.1, BA.2.83, XAW, and XBC. The NTD of SARS-CoV-2 is highly plastic, and
212 particularly prone to recurrent insertions and deletions which also show similar patterns of
213 convergence as substitutions in the RBD[43,44], including in Omicron lineages[45]. The phenotypic
214 impact of these recurrent convergent NTD insertions is yet to be properly characterised.

215 This seemingly rapid emergence of convergent immune escape mutations coincides
216 with a narrowing of the diversity of effective neutralising antibody responses with the
217 emergence of Omicron. Only a small subset of neutralising antibodies elicited by the ancestral
218 lineages of SARS-CoV-2 effectively cross-neutralise Omicron lineages[7,46]. Furthermore, after
219 an initial Omicron exposure, the neutralising antibody response is dominated by a subset of
220 reactivated memory B cells targeting epitopes conserved in the ancestral lineages[7,24]. This
221 includes an enrichment of 'class 3' antibodies specific for the epitopes outside the ACE2
222 binding site, but also public lineages targeting the RBD[7]. This narrowly focussed immune
223 pressure may be the driver of the extensive convergent evolution observed within these
224 epitopes[7,24].

It is unclear if such mutations will continue to accumulate over time at further, less immunodominant sites, or whether these mutations will slow down due to fitness costs associated with further mutations. Recent evidence also suggests that repeated Omicron exposure can result in the generation of *de novo* antibody responses, rather than just iterative boosting of ancestral-specific memory B cells[47]. This may lead to a re-broadening of antibody responses and less dramatic convergent evolution in the future.

As SARS-CoV-2 continues to evolve, continuously changing epistatic interactions created by RBD mutations provide favourable opportunities for novel antigenic change. For example, N460K appears to enhance human ACE2 binding which may then compensate for antigenic changes that reduce ACE2 affinity[7]. Linked to the changing epistatic fitness landscape, this continuous evolution also gives rise to a changing genetic landscape, facilitating the emergence of favourable amino acid substitutions[8,38] that have previously been rare due to the genetic context, in a manner somewhat similar to that described for seasonal influenza[48]. In this changing genetic landscape, an amino acid change that required a two-nucleotide change in the previous genetic context, may now be reached with a single mutation. Recent examples include emerging spike G339H, K478X, F486P and F490P variants, which are all arising across multiple branches of the Omicron phylogenetic tree.

The future of SARS-CoV-2 evolution

History of SARS-CoV-2 infection and vaccination were initially broadly similar for most individuals and, together with prevalent "public" antibody responses[49,50] – that is, highly similar antibodies in many individuals with shared genetic elements and modes of recognition – led to somewhat homogenous selection pressures on Spike.

As of mid-2023, XBB sublineages with F486P now dominate globally, and also represent the lineages with the fastest growth rates (for example XBB.1.5, XBB.1.16, XBB.1.9.1, XBB.1.9.2, XBB.2.3 and FE.1) (Figure 3B). However, descendants of BA.2.75, BA.5, BA.2.3.20, XAY and XBC continue to circulate at lower levels. Although these lineages contain mutations at many of the same sites in the RBD and NTD as XBB, often the exact amino acids involved differ (Figure 1). Therefore, it is possible subsequent immune responses elicited by these lineages may poorly cross-react with one another (particularly those derived from XBB,

BA.2.75 and BQ.1.1). During the first two years of the pandemic, co-circulation of SARS-CoV-2 variants was generally transient, with rapid lineage replacement[1]. However, in the future it is possible that several lineages with similar replicative fitness and enough antigenic distance from one another might co-circulate (in a manner similar to influenza B pre-2020), at least until a substantially fitter lineage or variant emerges.

To date, Omicron lineages have been the dominant SARS-CoV-2 variants for almost half the total duration of the pandemic[10]. Although all the lineages described here are genetically and antigenically distant from earlier Omicron lineages, when they have been tested *in vitro* or *in vivo* they continue to display comparable viral, epidemiological and clinical properties to their parent lineages, including a preference for endosomal cell entry, decreased syncytia formation, and decreased lower respiratory tract tropism compared to previous variants[25]. Despite the dominance of Omicron lineages throughout 2022 and into 2023, a novel variant that is not derived from Omicron might still emerge and replace current lineages, re-arising from a reservoir established prior to Omicron (for example, a chronically infected patient or animal). It is, however, unclear how frequently we should expect to see such events occur with so few previous examples. Since we cannot exclude the possibility of emergence in the future of an antigenically distinct lineage with higher pathogenicity than Omicron, it would be prudent to plan mitigation strategies, such as rapid vaccine updates, extra hospital capacity, stockpiles of antivirals and selected non-pharmaceutical interventions.

As well as new variants, it is also possible that continued evolution of Omicron could result in viruses with markedly different phenotypes. There is some evidence for modest variation in tissue tropism, protease preference, fusogenicity and pathogenicity among Omicron lineages[1]. Recombination of circulating Omicron lineages (particularly with Delta lineages, as seen in the complex recombinants) could also markedly alter viral properties; as has been shown for XD, an extinct recombinant between BA.1 and Delta that showed intermediate pathogenicity in a mouse model[28].

The evolution of a new variant via a saltatory-like variant emerging from a Delta genetic background is of particular concern, due to Delta's higher intrinsic severity than other VOC[51]. However, the intrinsic pathogenicity of Delta might change if it were to re-emerge.

Notably, a small number of Delta viruses continue to be sampled and sequenced worldwide, most with large numbers of private mutations, suggesting that there is a persistent reservoir of chronic Delta infections (as well as of other prior variants)[52]. Such sequences almost certainly represent ongoing chronic infections. In addition, less highly mutated pre-Omicron variants continue to be sampled from animal reservoirs, such as white-tailed deer or farmed mink, long after the variant has ceased circulation in the human population[53,54], leaving open the possibility for their reemergence from this source.

Genomic surveillance and analysis of SARS-CoV-2 should continue. SARS-CoV-2 is evolving according to well-understood mechanisms, but phenotypic and genetic directions of viral evolution are difficult to predict. Our experience of the SARS-CoV-2 pandemic highlights the importance of equitable global surveillance and sequencing capacity, which to date has been lacking[55]. It is vital that the lessons learned and sequencing capacity built are neither forgotten nor abandoned, but instead are maintained for SARS-CoV-2, as well as other respiratory viruses and viruses with zoonotic and pandemic potential.

As the pandemic has shown, the costs of such surveillance are trivial compared to the harm that viruses can cause.

Acknowledgements

We gratefully acknowledge all data contributors, i.e., the Authors and their Originating laboratories responsible for obtaining the specimens, and their Submitting laboratories for generating the genetic sequence and metadata and sharing via the GISAID Initiative[56], on which this research is based. The growth competition analysis from https://github.com/MurrellGroup/lineages[57] was performed on all available GISAID sequences, downloaded as a package on 2023-05-22.

The authors would also like to thank the Pango lineage designation committee and wider contributors for identifying and naming lineages, and the teams behind Pangolin, Nextstrain, covSPECTRUM, and UShER (particularly Angie Hinrichs) for developing tools that are used to identify and survey lineages. The authors would also like to thank Shay Fleishon for his valuable feedback and discussions. T.P.P is funded by the MRC funded G2P-UK National Virology Consortium (MR/W005611/1). O.G.P and C.R. are supported by the Oxford Martin

School. O.G.P. acknowledges assistance from Stephen Attwood, funded by Horizon Europe research and innovation programme under grant agreement No 871075 (ELIXIR-CONVERGE). D.J.S, K.S, and B.M. were supported by funding from SciLifeLab's Pandemic Laboratory Preparedness program (VC-2022-0028 and VC-2021-0033) and from the Erling Persson Foundation (ID: 2021 0125).

Conflicts of interest

A.O'T., A.R and O.G.P have undertaken consulting for AstraZeneca AB relating to the genetic diversity and classification of SARS-COV-2 lineages. D.J.S. also consults for AstraZeneca AB on the topic of monoclonal antibody therapeutics for COVID-19. At the time of final submission of this manuscript F. G. is now a contractor for Invivyd, Inc., a company that develops monoclonal antibodies to treat COVID-19. All other authors have no conflicts of interest to declare.

Author contributions

C.R., D.J.S., R.H., F.G., H.S., N.F., J.S., B.M. and T.P.P. researched data for article. C.R., D.J.S., R.H., F.G., H.S., N.F., J.S., A.O'T., A.R., O.G.P., C.R. substantially contributed to discussion of content. K.S., B.M. and T.P.P. performed updated analyses for the article. D.J.S. and T.P.P. wrote the first draft of the article. All authors reviewed and edited the manuscript before submission.

References

1 Carabelli, A. M. *et al.* SARS-CoV-2 variant biology: immune escape, transmission and fitness. *Nature Reviews Microbiology* **21**, 162-177, doi:10.1038/s41579-022-00841-7 (2023).

2 Harvey, W. T. *et al.* SARS-CoV-2 variants, spike mutations and immune escape. *Nature Reviews Microbiology* **19**, 409-424, doi:10.1038/s41579-021-00573-0 (2021).

3 Markov, P. V. *et al.* The evolution of SARS-CoV-2. *Nature Reviews Microbiology*, doi:10.1038/s41579-023-00878-2 (2023).

4 Schmidt, F. *et al.* High genetic barrier to SARS-CoV-2 polyclonal neutralizing antibody escape. *Nature* **600**, 512-516, doi:10.1038/s41586-021-04005-0 (2021).

5 Greaney, A. J. *et al.* Comprehensive mapping of mutations in the SARS-CoV-2 receptor-binding domain that affect recognition by polyclonal human plasma antibodies. *Cell Host & Microbe* **29**, 463-476.e466, doi:https://doi.org/10.1016/j.chom.2021.02.003 (2021).

6 Martin, D. P. *et al.* The emergence and ongoing convergent evolution of the SARS-CoV-2 N501Y lineages. *Cell* **184**, 5189-5200.e5187, doi:10.1016/j.cell.2021.09.003 (2021).

7 Cao, Y. *et al.* Imprinted SARS-CoV-2 humoral immunity induces convergent Omicron RBD evolution. *Nature* **614**, 521-529, doi:10.1038/s41586-022-05644-7 (2023).

8 Starr, T. N. *et al.* Deep Mutational Scanning of SARS-CoV-2 Receptor Binding Domain Reveals Constraints on Folding and ACE2 Binding. *Cell* **182**, 1295-1310.e1220, doi:https://doi.org/10.1016/j.cell.2020.08.012 (2020).

9 Zahradník, J. *et al.* SARS-CoV-2 variant prediction and antiviral drug design are enabled by RBD in vitro evolution. *Nature Microbiology* **6**, 1188-1198, doi:10.1038/s41564-021-00954-4 (2021).

10 Viana, R. *et al.* Rapid epidemic expansion of the SARS-CoV-2 Omicron variant in southern Africa. *Nature*, doi:10.1038/s41586-022-04411-y (2022).

11 Tegally, H. *et al.* Detection of a SARS-CoV-2 variant of concern in South Africa. *Nature* **592**, 438-443, doi:10.1038/s41586-021-03402-9 (2021).

12 Faria, N. R. *et al.* Genomics and epidemiology of the P.1 SARS-CoV-2 lineage in Manaus, Brazil. *Science* **372**, 815-821, doi:doi:10.1126/science.abh2644 (2021).

13 Rambaut, A. *et al.* Preliminary genomic characterisation of an emergent SARS-CoV-2 lineage in the UK defined by a novel set of spike mutations. *Virological.org* (2020).

14 Hill, V. *et al.* The origins and molecular evolution of SARS-CoV-2 lineage B.1.1.7 in the UK. *Virus Evolution* **8**, doi:10.1093/ve/veac080 (2022).

15 Ruis, C. *et al.* Preadaptation of pandemic GII.4 noroviruses in unsampled virus reservoirs years before emergence. *Virus Evolution* **6**, doi:10.1093/ve/veaa067 (2020).

16 Kemp, S. A. *et al.* SARS-CoV-2 evolution during treatment of chronic infection. *Nature* **592**, 277-282, doi:10.1038/s41586-021-03291-y (2021).

17 Wilkinson, S. A. J. *et al.* Recurrent SARS-CoV-2 mutations in immunodeficient patients. *Virus Evolution* **8**, doi:10.1093/ve/veac050 (2022).

18 Harari, S. *et al.* Drivers of adaptive evolution during chronic SARS-CoV-2 infections. *Nature Medicine*, doi:10.1038/s41591-022-01882-4 (2022).

19 Mallapaty, S. Where did Omicron come from? Three key theories. *Nature* **602**, 26-28, doi:10.1038/d41586-022-00215-2 (2022).

20 Tegally, H. *et al.* Emergence of SARS-CoV-2 Omicron lineages BA.4 and BA.5 in South Africa. *Nature Medicine*, doi:10.1038/s41591-022-01911-2 (2022).

21 Karyakarte, R. P. *et al.* An Early and Preliminary Assessment of the Clinical Severity of the Emerging SARS-CoV-2 Omicron Variants in Maharashtra, India. *Cureus* **14**, e31352, doi:10.7759/cureus.31352 (2022).

22 Japanese National Institute of infectious Diseases (NIID). 感染・伝播性の増加や抗原性の変化が懸念される 新型コロナウイルス（SARS-CoV-2）の変異株について （第20報）. (www.niid.go.jp, 2022).

23 Hisner, R. *2nd-Generation BA.2 Saltation Lineage, >30 spike mutations (3 seq, 2 countries, Aug 14) #2183*, <https://github.com/cov-lineages/pango-designation/issues/2183> (2023).

24 Dijokaite-Guraliuc, A. *et al.* Rapid escape of new SARS-CoV-2 Omicron variants from BA.2-directed antibody responses. *Cell Rep* **42**, 112271, doi:10.1016/j.celrep.2023.112271 (2023).

25 UKHSA. SARS-CoV-2 variants of concern and variants under investigation in England: technical briefing 49. (2023).

26 Jackson, B. *et al.* Generation and transmission of interlineage recombinants in the SARS-CoV-2 pandemic. *Cell* **184**, 5179-5188.e5178, doi:10.1016/j.cell.2021.08.014 (2021).

27 Gutierrez, B. *et al.* Emergence and widespread circulation of a recombinant SARS-CoV-2 lineage in North America. *Cell Host Microbe* **30**, 1112-1123.e1113, doi:10.1016/j.chom.2022.06.010 (2022).

28 Simon-Loriere, E. *et al.* Rapid characterization of a Delta-Omicron SARS-CoV-2 recombinant detected in Europe. *Research Square*, doi:10.21203/rs.3.rs-1502293/v1 (2022).

29 Pybus, O. Pango Lineage Nomenclature: provisional rules for naming recombinant lineages. *Virological.org* (2021).

30 Sheward, D. J. *et al.* Omicron sublineage BA.2.75.2 exhibits extensive escape from neutralising antibodies. *The Lancet Infectious Diseases* **22**, 1538-1540, doi:10.1016/S1473-3099(22)00663-6 (2022).

31 Akerman, A. *et al.* Emergence and antibody evasion of BQ, BA.2.75 and SARS-CoV-2 recombinant sub-lineages in the face of maturing antibody breadth at the population level. *eBioMedicine* **90**, doi:10.1016/j.ebiom.2023.104545 (2023).

32 Silcn. *XBB.1*/BA.2.75*/XBB.1* recombinant with S:F486P (8 seq, Malaysia) #1532*, <https://github.com/cov-lineages/pango-designation/issues/1532> (2023).

33 (NGS, N. f. G. S. i. S. A. & SA). SARS-CoV 2 Sequencing Update 15 July 2022. (NGS-SA report, 2022).

34 Pangilinan, E. A. R. *et al.* Analysis of SARS-CoV-2 Recombinant Lineages XBC and XBC.1 in the Philippines and Evidence for Delta-Omicron Co-infection as a Potential Origin. *bioRxiv*, 2023.2004.2012.534029, doi:10.1101/2023.04.12.534029 (2023).

35 Chaguza, C. *et al.* Accelerated SARS-CoV-2 intrahost evolution leading to distinct genotypes during chronic infection. *medRxiv*, 2022.2006.2029.22276868, doi:10.1101/2022.06.29.22276868 (2022).

36 Smith, D. J. *et al.* Mapping the Antigenic and Genetic Evolution of Influenza Virus. *Science* **305**, 371-376, doi:doi:10.1126/science.1097211 (2004).

37 Eguia, R. T. *et al.* A human coronavirus evolves antigenically to escape antibody immunity. *PLoS Pathog* **17**, e1009453, doi:10.1371/journal.ppat.1009453 (2021).

38 Moulana, A. *et al.* Compensatory epistasis maintains ACE2 affinity in SARS-CoV-2 Omicron BA.1. *Nature Communications* **13**, 7011, doi:10.1038/s41467-022-34506-z (2022).

39 Yue, C. *et al.* ACE2 binding and antibody evasion in enhanced transmissibility of XBB.1.5. *Lancet Infect Dis*, doi:10.1016/s1473-3099(23)00010-5 (2023).

40 Focosi, D., Quiroga, R., McConnell, S., Johnson, M. C. & Casadevall, A. Convergent Evolution in SARS-CoV-2 Spike Creates a Variant Soup from Which New COVID-19 Waves Emerge. *International Journal of Molecular Sciences* **24**, 2264 (2023).

41 Starr, T. N. *et al.* Deep mutational scans for ACE2 binding, RBD expression, and antibody escape in the SARS-CoV-2 Omicron BA.1 and BA.2 receptor-binding domains. *PLoS Pathog* **18**, e1010951, doi:10.1371/journal.ppat.1010951 (2022).

42 Cox, M. *et al.* SARS-CoV-2 variant evasion of monoclonal antibodies based on in vitro studies. *Nature Reviews Microbiology* **21**, 112-124, doi:10.1038/s41579-022-00809-7 (2023).

43 McCallum, M. *et al.* N-terminal domain antigenic mapping reveals a site of vulnerability for SARS-CoV-2. *Cell* **184**, 2332-2347.e2316, doi:10.1016/j.cell.2021.03.028 (2021).

44 Gerdol, M., Dishnica, K. & Giorgetti, A. Emergence of a recurrent insertion in the N-terminal domain of the SARS-CoV-2 spike glycoprotein. *Virus Res* **310**, 198674, doi:10.1016/j.virusres.2022.198674 (2022).

45 Greco, S. & Gerdol, M. Independent acquisition of short insertions at the RIR1 site in the spike N-terminal domain of the SARS-CoV-2 BA.2 lineage. *Transboundary and Emerging Diseases* **69**, e3408-e3415, doi:https://doi.org/10.1111/tbed.14672 (2022).

46 Cao, Y. *et al.* Omicron escapes the majority of existing SARS-CoV-2 neutralizing antibodies. *Nature* **602**, 657-663, doi:10.1038/s41586-021-04385-3 (2022).

47 Yisimayi, A. *et al.* Repeated Omicron infection alleviates SARS-CoV-2 immune imprinting. *bioRxiv*, 2023.2005.2001.538516, doi:10.1101/2023.05.01.538516 (2023).

48 Koelle, K., Cobey, S., Grenfell, B. & Pascual, M. Epochal evolution shapes the phylodynamics of interpandemic influenza A (H3N2) in humans. *Science* **314**, 1898-1903, doi:10.1126/science.1132745 (2006).

49 Barnes, C. O. *et al.* SARS-CoV-2 neutralizing antibody structures inform therapeutic strategies. *Nature* **588**, 682-687, doi:10.1038/s41586-020-2852-1 (2020).

50 Wang, Y. *et al.* A large-scale systematic survey reveals recurring molecular features of public antibody responses to SARS-CoV-2. *Immunity* **55**, 1105-1117.e1104, doi:10.1016/j.immuni.2022.03.019 (2022).
51 Twohig, K. A. *et al.* Hospital admission and emergency care attendance risk for SARS-CoV-2 delta (B.1.617.2) compared with alpha (B.1.1.7) variants of concern: a cohort study. *The Lancet Infectious Diseases* **22**, 35-42, doi:10.1016/S1473-3099(21)00475-8 (2022).
52 Harari, S., Miller, D., Fleishon, S., Burstein, D. & Stern, A. Using big sequencing data to identify chronic SARS-Coronavirus-2 infections. *bioRxiv*, 2023.2007.2016.549184, doi:10.1101/2023.07.16.549184 (2023).
53 Domańska-Blicharz, K. *et al.* Cryptic SARS-CoV-2 lineage identified on two mink farms as a possible result of long-term undetected circulation in an unknown animal reservoir, Poland, November 2022 to January 2023. *Eurosurveillance* **28**, 2300188, doi:doi:https://doi.org/10.2807/1560-7917.ES.2023.28.16.2300188 (2023).
54 Pickering, B. *et al.* Divergent SARS-CoV-2 variant emerges in white-tailed deer with deer-to-human transmission. *Nature Microbiology* **7**, 2011-2024, doi:10.1038/s41564-022-01268-9 (2022).
55 Brito, A. F. *et al.* Global disparities in SARS-CoV-2 genomic surveillance. *Nature Communications* **13**, 7003, doi:10.1038/s41467-022-33713-y (2022).
56 Shu, Y. & McCauley, J. GISAID: Global initiative on sharing all influenza data – from vision to reality. *Eurosurveillance* **22**, 30494, doi:doi:https://doi.org/10.2807/1560-7917.ES.2017.22.13.30494 (2017).
57 Murrell, B. *SARS-CoV-2 Lineage Competition*, <https://github.com/MurrellGroup/lineages> (2023).
58 Lacek, K. A. *et al.* SARS-CoV-2 Delta-Omicron Recombinant Viruses, United States. *Emerging Infectious Diseases* **28** (2022).

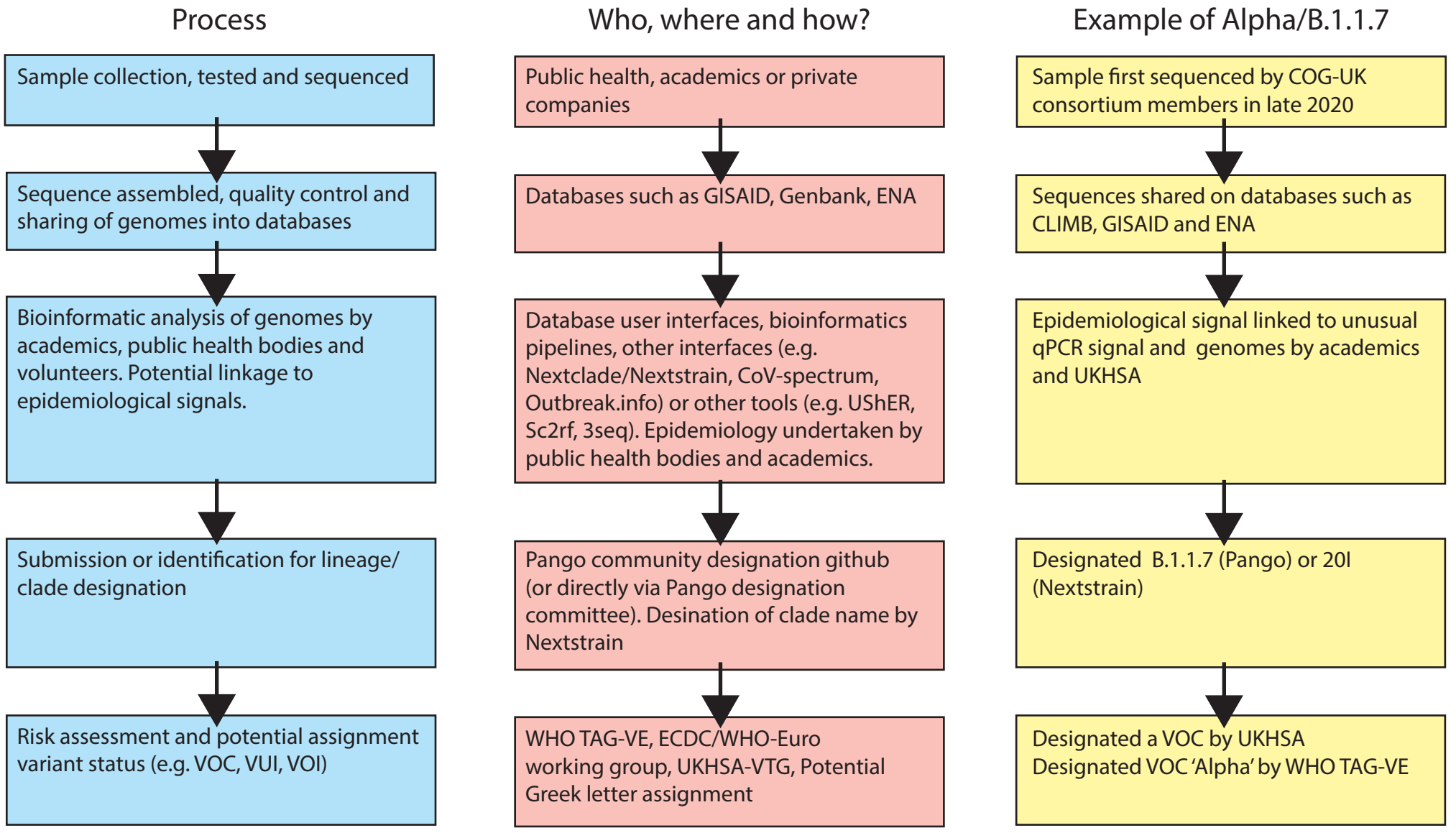
Box 1. How new SARS-CoV-2 variants are sampled, identified and categorised. With record levels of sequenced viral genomes being generated during the COVID-19 pandemic it became important to develop and implement new tools, protocols and nomenclature systems to aid in interpretation of the virus evolution and epidemiology. These have come from a range of different sources including academia, public health bodies, private companies and even citizen science.

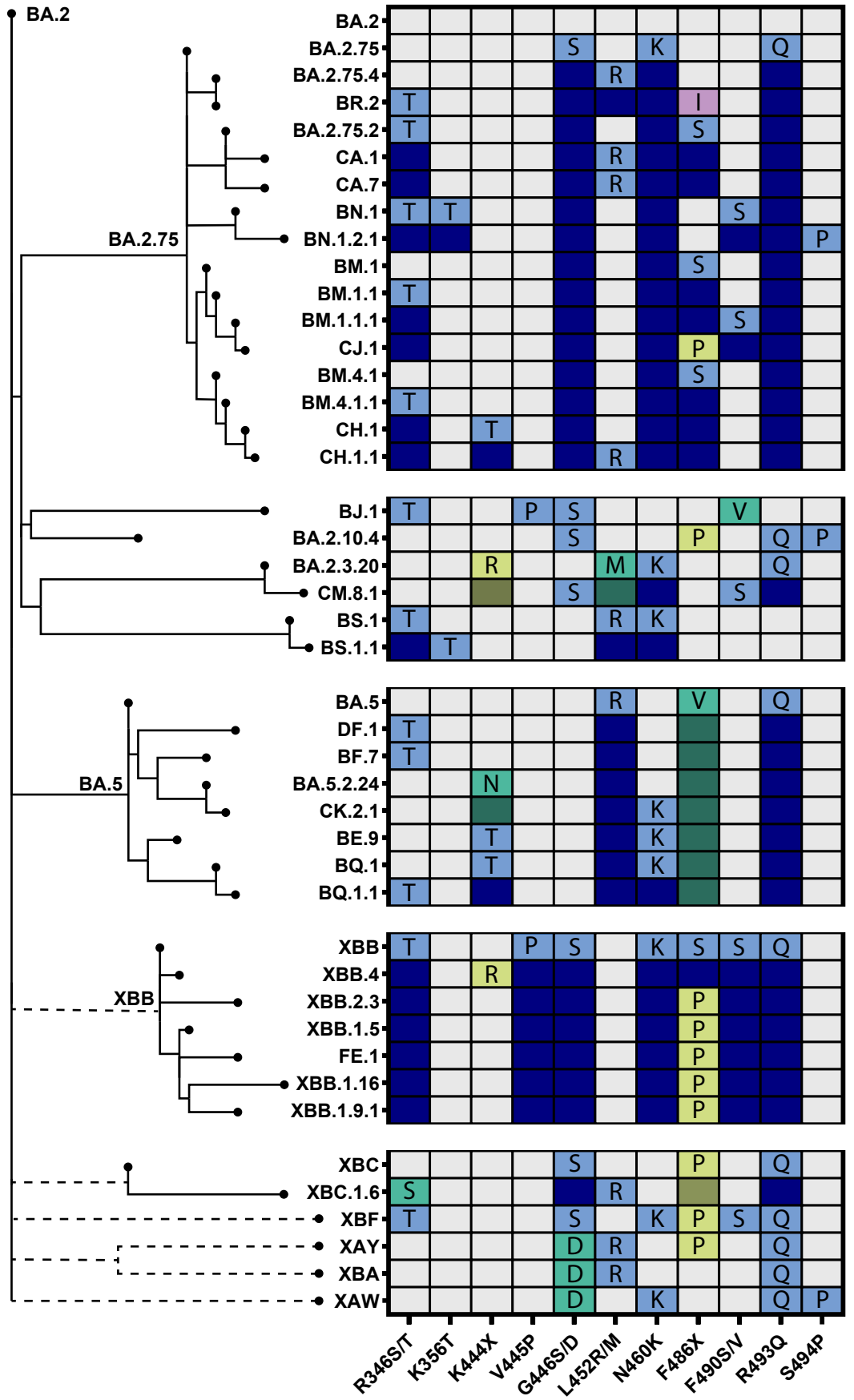
Figure 1. Phylogenetic relatedness and convergent evolution in contemporary (as of July 2023) Omicron descendent lineages. A simplified phylogenetic tree, modified manually from the Nextclade curated tree, showing a selection of fast growing or large lineages[30] that exhibit convergent molecular evolution. The table shows amino acid positions in the spike protein that exhibit the most prominent convergent evolutionary patterns. Grey boxes indicate that the amino acid at the site is the same as that in the root ancestor of BA.2. Coloured boxes indicate amino acid changes at that site (with different colours arbitrarily showing different amino acids). Light shading with a letter inside indicates new mutations on the branch, dark

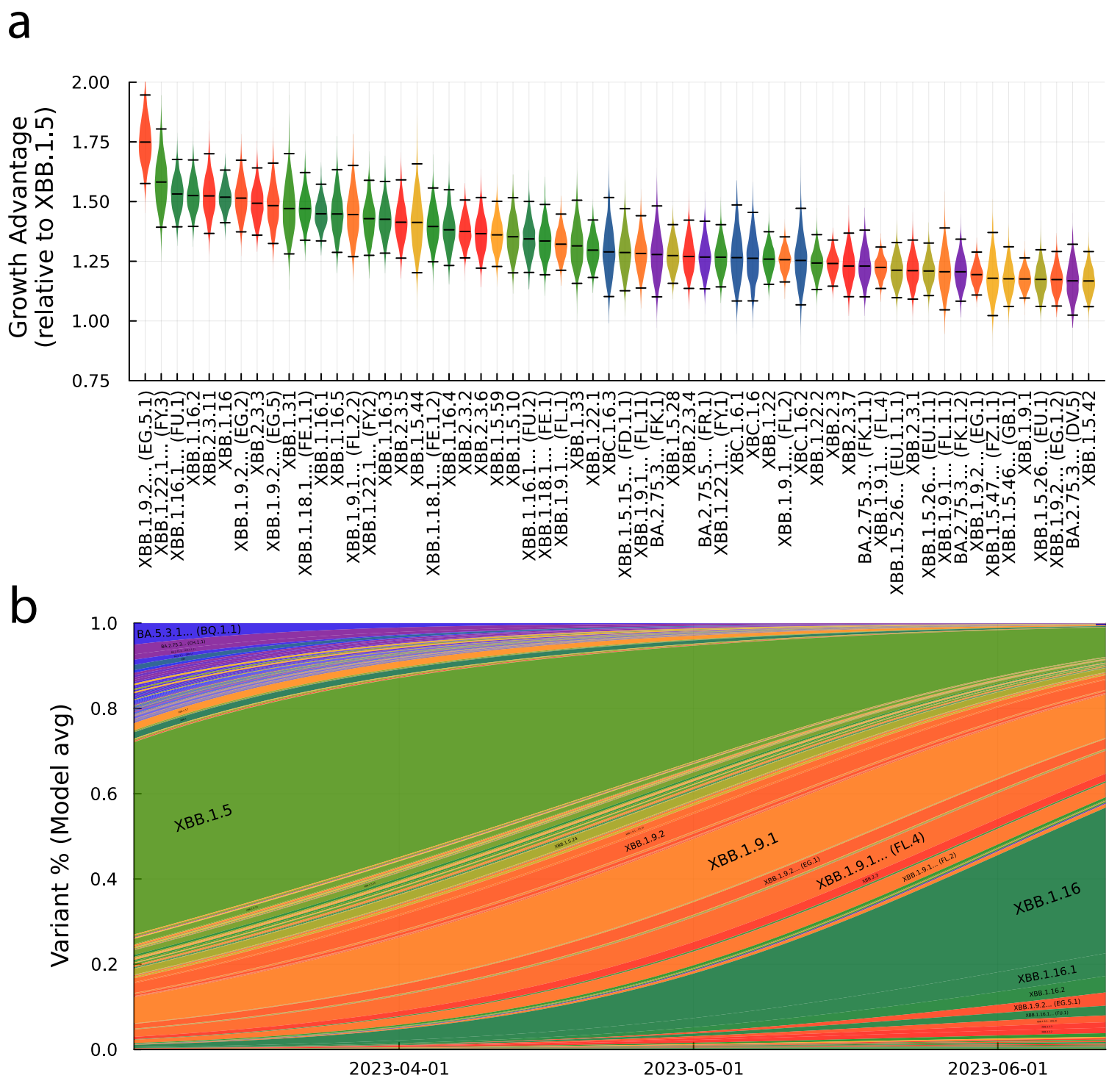
shading indicates mutations inherited from a parental lineage. Branches leading into recombinant lineages shown as dotted lines. Amino acid X on the X-axis labels indicates any amino acid substitution at that site.

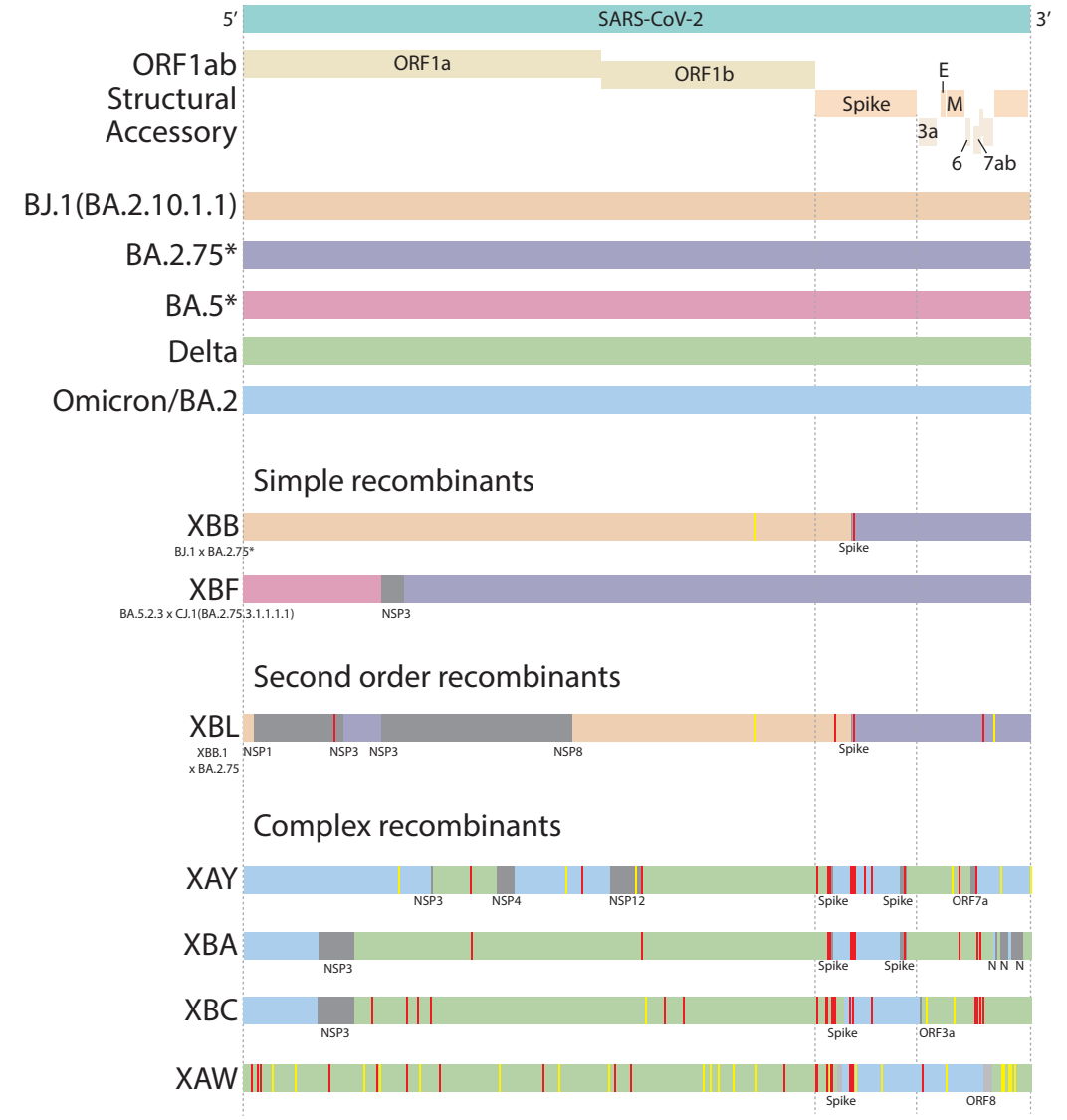
Figure 2. Relative growth rates and variant proportions, as of 22nd May 2023. (A) 60 fastest growing SARS-CoV-2 lineages relative to XBB.1.5. Lineage colours are grouped by relatedness. (B) Expected variant proportions under growth rates inferred from data sampled globally. Lineage competition was modelled using a Bayesian multinomial regression approach described previously[30,31] and maintained online[57].

Figure 3. Genome schematics of the contemporary recombinant lineages. Different colours indicate different parts of the genome from each parent. Grey areas indicate ambiguous areas that most likely contain the breakpoint. Red lines indicate non-synonymous private mutations, while yellow lines indicate synonymous private mutations. Recombinant break point analysis was performed as performed previous[58], manually using presence of non-convergent private mutations from putative parental lineages (identified as those sharing the majority of private mutations in the relevant genomic region).


Process
Who, where and how?
Example of Alpha/B.1.1.7
Sample collection, tested and sequenced
Public health, academics or private companies
Sample first sequenced by COG-UK consortium members in late 2020
Sequence assembled, quality control and sharing of genomes into databases
Databases such as GISAID, Genbank, ENA
Sequences shared on databases such as CLIMB, GISAID and ENA
Bioinformatic analysis of genomes by academics, public health bodies and volunteers. Potential linkage to epidemiological signals.
Database user interfaces, bioinformatics pipelines, other interfaces (e.g. Nextclade/Nextstrain, CoV-spectrum, Outbreak.info) or other tools (e.g. UShER, Sc2rf, 3seq). Epidemiology undertaken by public health bodies and academics.
Epidemiological signal linked to unusual qPCR signal and genomes by academics and UKHSA
Submission or identification for lineage/ clade designation
Pango community designation github (or directly via Pango designation committee). Desination of clade name by Nextstrain
Designated B.1.1.7 (Pango) or 20I (Nextstrain)
Risk assessment and potential assignment variant status (e.g. VOC, VUI, VOI)
WHO TAG-VE, ECDC/WHO-Euro working group, UKHSA-VTG, Potential Greek letter assignment
Designated a VOC by UKHSA
Designated VOC 'Alpha' by WHO TAG-VE


BA.2
BA.2.75
BA.5
XBB
BA.2
BA.2.75
BR.2
BA.2.75.2
CA.1
CA.7
BN.1
BN.1.2.1
BM.1
BM.1.1
BM.1.1.1
CJ.1
BM.4.1
BM.4.1.1
CH.1
CH.1.1
BJ.1
BA.2.10.4
BA.2.3.20
CM.8.1
BS.1
BS.1.1
BA.5
DF.1
BF.7
BA.5.2.24
CK.2.1
BE.9
BQ.1
BQ.1.1
XBB
XBB.4
XBB.2.3
XBB.1.5
FE.1
XBB.1.16
XBB.1.9.1
XBC
XBC.1.6
XBF
XAY
XBA
XAW
R346S/T
K356T
K444X
V445P
G446S/D
L452R/M
N460K
F486X
F490S/V
R493Q
S494P


a
Growth Advantage
(relative to XBB.1.5)
2.00
1.75
1.50
1.25
1.00
0.75
XBB.1.9.2... (EG.5.1)
XBB.1.22.1... (FY.3)
XBB.1.16.1... (FU.1)
XBB.1.16.2
XBB.2.3.11
XBB.1.16
XBB.1.9.2... (EG.2)
XBB.2.3.3
XBB.1.9.2... (EG.5)
XBB.1.31
XBB.1.18.1... (FE.1.1)
XBB.1.16.1
XBB.1.16.5
XBB.1.9.1... (FL.2.2)
XBB.1.22.1... (FY.2)
XBB.1.16.3
XBB.2.3.5
XBB.1.5.44
XBB.1.18.1... (FE.1.2)
XBB.1.16.4
XBB.2.3.2
XBB.2.3.6
XBB.1.5.59
XBB.1.5.10
XBB.1.16.1... (FU.2)
XBB.1.18.1... (FE.1)
XBB.1.9.1... (FL.1)
XBB.1.33
XBB.1.22.1
XBC.1.6.3
XBB.1.5.15... (FD.1.1)
XBB.1.9.1... (FL.11)
BA.2.75.3... (FK.1)
XBB.1.5.28
XBB.2.3.4
BA.2.75.5... (FR.1)
XBB.1.22.1... (FY.1)
XBC.1.6.1
XBC.1.6
XBB.1.22
XBB.1.9.1... (FL.2)
XBC.1.6.2
XBB.1.22.2
XBB.2.3
XBB.2.3.7
BA.2.75.3... (FK.1.1)
XBB.1.9.1... (FL.4)
XBB.1.5.26... (EU.1.1.1)
XBB.2.3.1
XBB.1.5.26... (EU.1.1)
XBB.1.9.1... (FL.1.1)
BA.2.75.3... (EG.1)
XBB.1.9.2... (FZ.1.1)
XBB.1.5.46... (GB.1)
XBB.1.9.1
XBB.1.5.26... (EU.1)
XBB.1.9.2... (EG.1.2)
BA.2.75.3... (DV.5)
XBB.1.5.42
b
Variant % (Model avg)
1.0
0.8
0.6
0.4
0.2
0.0
BA.5.3.1... (BQ.1.1)
XBB.1.5
XBB.1.9.2
XBB.1.9.1
XBB.1.9.1... (FL.4)
XBB.1.16
XBB.1.16.1
XBB.1.16.2
XBB.1.9.2... (EG.5.1)
2023-04-01
2023-05-01
2023-06-01


5′
SARS-CoV-2
3′
ORF1ab
Structural
Accessory
ORF1a
ORF1b
Spike
E
M
3a
6
7ab
BJ.1(BA.2.10.1.1)
BA.2.75*
BA.5*
Delta
Omicron/BA.2
Simple recombinants
XBB
BJ.1 x BA.2.75*
Spike
XBF
BA.5.2.3 x CJ.1(BA.2.75.3.1.1.1.1)
NSP3
Second order recombinants
XBL
XBB.1
x BA.2.75
NSP1
NSP3
NSP3
NSP8
Spike
Complex recombinants
XAY
NSP3
NSP4
NSP12
Spike
Spike
ORF7a
XBA
NSP3
Spike
Spike
N N N
XBC
NSP3
Spike
ORF3a
XAW
Spike
ORF8