Adaptive advantage of deletion repair in the N-terminal domain of the SARS-CoV-2 spike protein in variants of concern

Miguel Álvarez-Herrera[a], Paula Ruiz-Rodriguez[a], Beatriz Navarro-Domínguez[ab], Joao Zulaica[a], Brayan Grau[a], María Alma Bracho[ac], Manuel Guerreiro[de], Cristóbal Aguilar Gallardo[e], Fernando González-Candelas[ac], Iñaki Comas[cf], Ron Geller[a], Mireia Coscollá[a]*

[a]*Institute for Integrative Systems Biology (I^{2}SysBio, University of Valencia-CSIC), FISABIO Joint Research Unit "Infection and Public Health", Paterna, Spain;* [b]*Universtiy of Granada, Department of Genetics, Granada, Spain;* [c]*CIBER in Epidemiology and Public Health (CIBERESP), Madrid, Spain;* [d]*Haematology Department, La Fe University and Polytechnic Hospital, Valencia, Spain;* [e]*La Fe Health Research Institute (IIS-La Fe), Valencia, Spain;* [f]*Institute of Biomedicine of Valencia (IBV-CSIC), Valencia, Spain*

* Email: mireia.coscolla@csic.es. Address: Institute for Integrative Systems Biology, C/ Catedrático Agustín Escardino 9, Building 4, 46980 Paterna, Spain

Keywords

SARS-CoV-2, COVID-19, spike, transmission, deletion, fusogenicity, antibody neutralization, thermal stability

Abstract

Mutations within the N-terminal domain (NTD) of the spike (S) protein play a pivotal role in the emergence of successful SARS-CoV-2 viral lineages. This study investigates the influence on viral success of novel combinations of NTD lineage-defining mutations found in the Alpha, Delta, and Omicron variants. We performed comparative genomics of more than 10 million public SARS-CoV-2 samples to decipher the transmission success of different combinations of NTD markers. Additionally, we characterized the viral phenotype of such markers in a surrogate *in vitro* system. Alpha viruses bearing repaired deletions S:ΔH69/V70 and S:ΔY144 in Alpha background were associated with increased transmission relative to other combinations of NTD markers. After the emergence of the Omicron BA.1 lineage, Alpha viruses harbouring both repaired deletions still showed increased transmission compared to their BA.1 analogues. Moreover, repaired deletions were more frequently observed among older individuals infected with Alpha, but not with BA.1. *In vitro* biological characterization of Omicron BA.1 spike deletion repair patterns also revealed substantial differences with Alpha. In BA.1, S:ΔV143/Y145 repair enhanced fusogenicity and susceptibility to neutralization by vaccinated individuals' sera. In contrast, the S:ΔH69/V70 repair did not significantly alter these traits but reduced viral infectivity. Simultaneous repair of both deletions led to lower fusogenicity. These findings highlight the intricate genotype-phenotype landscape of the spike NTD in SARS-CoV-2, which impacts viral biology, transmission efficiency, and susceptibility to neutralization. Overall, this study advances our understanding of SARS-CoV-2 evolution, carrying implications for public health and future research.

Introduction

As we navigate the constantly shifting landscape of the COVID-19 pandemic, the remarkable potential of SARS-CoV-2 for genetic adaptation has taken centre stage. Global turnover of SARS-CoV-2 lineages happened several times, with three variants of concern (VOCs) displacing the previous predominant lineages just in 2021: first, Alpha (B.1.1.7 and Q sublineages), then Delta (B.1.617.2 and AY sublineages), and lastly, Omicron (B.1.1.529 and BA sublineages, among others) [1]. This successive replacement of lineages across the pandemic suggests that the newer lineages had a higher adaptive value than the previous ones [2,3], and thus, these three lineages are supposed to carry mutations associated with higher transmissibility and/or immune evasion.

The spike (S) protein —a type I transmembrane N-linked glycosylated protein of 150–200 kDa— is a hotspot for mutations with high adaptive value. Spike proteins are located on the surface of SARS-CoV-2, and their main role is to mediate viral cell entry [4]. The spike protein forms a homotrimer, which is cleaved post-transcriptionally into two subunits: S1 and S2. The S1 consists of the amino or N-terminal domain (NTD) and the receptor-binding domain (RBD), and it is responsible for binding to the host cell-surface receptor, ACE2. The S2 subunit includes the trimeric core of the protein and is responsible for membrane fusion [5,6]. Therefore, some amino acid changes in the S protein may confer an advantage for transmission, considering its role in mediating viral cell entry. Additionally, most antibodies target sites either in the NTD or RBD, and therefore, mutations in these regions may enhance immune escape.

In Alpha lineages, deletions in the NTD (S:ΔH69/V70 and S:ΔY144) are associated with antibody escape [7,8] and increased infectivity [9]. Acquisition of deletions in the NTD of the spike glycoprotein during long-term infections of

immunocompromised patients has been reported and identified as an evolutionary pattern defined by recurrent deletions that alter defined antibody epitopes [10,11]. Additionally, deletions may play a decisive role in SARS-CoV-2 adaptive evolution, particularly on deletion-tolerant genome regions such as the S gene, as they can hardly be corrected by the proofreading activity of its RNA-dependent RNA polymerase (RdRP) [12–14]. Indeed, the NTD has been extensively impacted by deletions, which have arisen multiple times in different variants, including Alpha, Delta, and Omicron. For example, different deletions are observed in Delta (S:ΔE156/F157-R158G), Alpha (S:ΔY144) and BA.1 (S:ΔV143/Y145) variants, all mapping to the same surface indicating a convergent function [15,16]. All these mutations are within the NTD site targeted by most anti-NTD neutralizing antibodies [8]. Despite not being found in BA.2, its descendant lineage BJ.1 seems to have independently acquired S:ΔY144 (see cov-lineages/pango-designation issues #915 and #922 on GitHub). Then, it was passed on to XBB viruses through recombination with a BA.2-descended lineage, where it is associated with increased immune escape regardless of the vaccination status or infection history [17–19]. Subsequently, in January 2023, the newly emerging XBB.1.5 lineage was shown to display an increased receptor-binding affinity and infectivity with respect to its parental lineage [20,21].

During the takeover of the Alpha variant by the Delta variant, we became intrigued by the lack of overlap in NTD mutations between them. If mutations in the NTD increased immune escape without compromising binding to ACE2, one might expect that mutations in NTD have a cumulative (“the more the better”) effect. However, that would not be the case if epistatic interactions between sites prevent the fixation of particular mutations in different genomic and genetic backgrounds. Our primary objective in this study is to investigate the effect of NTD deletion repair in

variants of concern. We focus on examining the differences in success between variants of concern as a function of different mutational patterns in the NTD, including recurrent deletion repair events occurring in distinct independent lineages. Our findings suggest a non-linear effect of specific deletion repairs on viral phenotype, highlighting the importance of examining the genomic context of SARS-CoV-2 mutations.

Materials and Methods

Data retrieval and preprocessing

Screening Alpha and Delta combinations of NTD markers included 7,133,237 available SARS-CoV-2 sequences fetched from the GISAID [22] EpiCoV database on January 16, 2022. The GISAID EPI_SET is available at DOI: 10.55876/gis8.230802fh. We ran Nextalign CLI v1.9.0 [23] with default scoring settings to obtain aligned genomes and peptides corresponding to each gene. Genomes from non-human hosts were discarded. Then, we filtered out genomes containing more than 5% ambiguous bases and 1,000 gaps, or at least one indeterminate position among the lineage-defining sites of the spike protein. 6,589,393 genomes passed these filters. An additional filtering step implemented in Nextclade CLI v1.7.0 [24] was used to flag and remove false positive mixtures possibly caused by sample contamination or coinfections, considering that we were exploring combinatorial variation during the time of co-existence of highly successful variants. The same dataset was used to screen for patterns of deletion repair in the Alpha variant, omitting the Nextclade filtering step as it was not deemed critical when not searching for variant mixtures. We performed an analogous search of deletion-missing sequences assigned to lineage BA.1 out of 11,334,504 samples (10,353,158 after quality-filtering), fetched on June 15, 2022. The GISAID EPI_SET is available at DOI: 10.55876/gis8.230801ex.

The classification of Alpha-Delta combinatorial haplotypes was based on the sequence correspondence of protein segments to the canonical variants. We defined three blocks within the N-terminus of the S protein that included all NTD lineage-defining mutations of the Alpha (S:ΔH69/V70 and S:ΔY144) and Delta (S:T19R and S:ΔE156/F157 + S:R158G) variants, each encompassing contiguous lineage-defining sites (S:16, S:69/70 + S:144 and S:156/158; see Figure 1B). We considered the remaining portion of the protein as background. Alpha and Delta backgrounds include the mutations in Figure 1A. Site S:142 was not considered in this analysis due to a known technical sequencing artifact in variant Delta [25]. See Supplemental Note 1 for further details about haplotype naming. This survey included sixteen haplotypes, including the canonical Alpha and Delta spike protein sequences and fourteen combinatorial haplotypes (I-XIV).

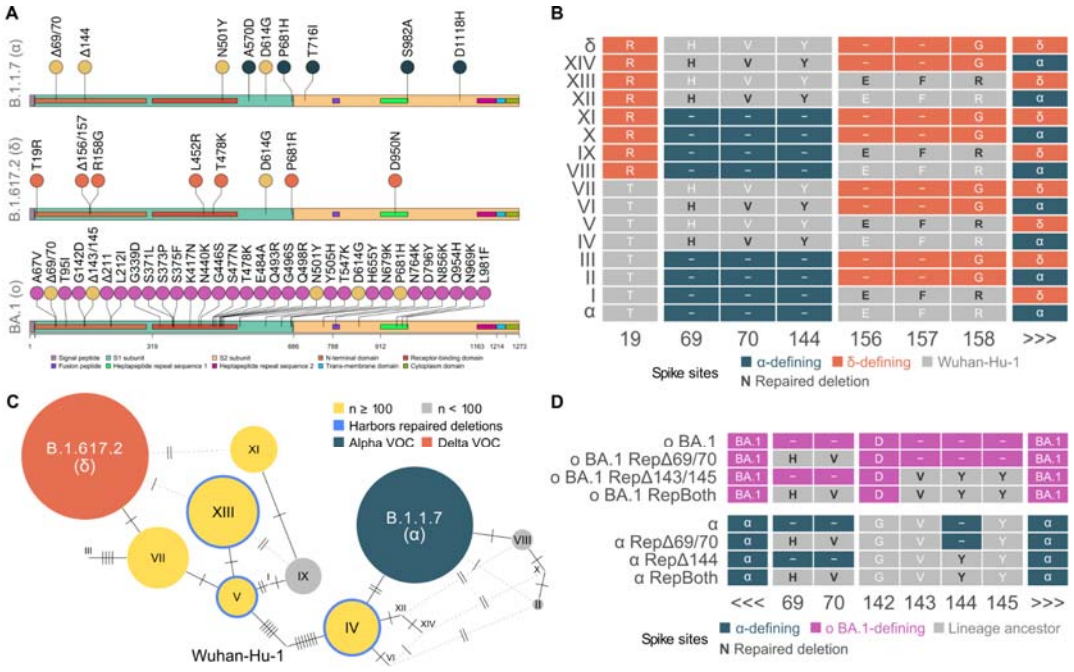

Figure 1. Combinatorial standpoint of the study. (A) Defining mutations of the B.1.1.7 (Alpha), B.1.617.2 (Delta) and BA.1 (Omicron) SARS-CoV-2 lineages on the spike protein. (B) Spike protein lineage-defining sites of the fourteen Alpha-Delta combinatorial haplotypes (I-XIV) and two variants of concern. (C) Haplotype network of the most abundant combinatorial haplotypes (see Table 1) and the B.1.1.7 and B.1.617.2 lineages, based on the haplotype group-level consensus of the S gene. Solid edges depict the minimum spanning tree of the graph. Each stroke on the edges stands for one nucleotide site that changes between the connected nodes. Node size represents the number of observations in log-scale. (D) Spike protein lineage-defining sites of samples that repair NTD deletions in common on Alpha and Omicron BA.1 background. Sites in grey are unaltered with respect to each lineage ancestral sequence.

Table 1. Description of five Alpha-Delta combinatorial haplotypes with at least 100 observations. Values marked with † were obtained with the dataset reduction strategy with CD-HIT.

Haplotype	Description	Observations	Transmission clusters	Minimum independent emergences
IV	Alpha + repaired S:ΔH69/V70 & S:ΔY144	736	72 78†	576 598†
V	Delta + S:R19T reversion & repaired S:ΔE156/F157-R158G	117	6	104
VII	Delta + S:R19T reversion	1,666	146	1,121
XI	Delta + S:ΔH69/V70 & S:ΔY144	304	19	267
XIII	Delta + repaired S:ΔE156/F157-R158G	4,781	431	2,880

Analogously, the survey of NTD deletion repair haplotypes in Alpha and Omicron BA.1 backgrounds tracked the presence or absence of specific deletions in sequences assigned to lineages B.1.1.7 (Alpha; deletions S:H69/V70 and S:ΔY144) and BA.1 (Omicron; deletions S:H69/V70 and S:ΔV143/Y145), or any of their sublineages. For more in-depth considerations about S:ΔV143/Y145 repair in lineage BA.1, see Supplemental Note 2.

After data cleaning and homogenization of the retrieved sample metadata, we were able to obtain the host age for approximately 50% of the analyzed samples of both datasets. We used R v4.1.2 [26] along with *tidyverse* v1.3.1 [27] to conduct, manage and visualize these analyses. Lollipop plots were generated using *trackViewer* v1.30.0 [28]. Haplotype networks with abundance weights of the S gene consensus sequences of the combinatorial haplotypes were built using the randomized minimum spanning tree (RMST) algorithm [29] implemented in *pegas* v1.1 [30]. We considered insertion or deletion blocks as a single mutational event. Synthetic sequences were built to account for unsampled haplotypes.

Number of emergences and transmissions

To infer the success of each combinatorial haplotype, we assessed the number of

minimum emergence events and whether these originated subsequent transmission events. To overcome the challenging computational demands of working with global-scale sequence datasets and, we devised a double approach, based on two distinct optimality criteria. First, we placed genomes matching combinatorial haplotypes (I-XIV) on a comprehensive phylogeny of public sequences up to the date of each survey [31] under a maximum parsimony criterion, using UShER v0.5.3 [32]. Second, we used an additional phylogenetic inference method that considered an evolutionary process model while keeping the phylogenetic context of our datasets, to enhance the reliability of our results. S gene clustering based on short word filtering was performed with CD-HIT-EST, bundled in the CD-HIT suite v4.8.1 [33], with a word size of 8 and a minimum sequence identity threshold equivalent to 5 amino acid changes in the spike protein. For any given haplotype, sequences within clusters that included at least one of the haplotype-assigned samples were selected as members of the reduced dataset for the corresponding haplotype. Then, we inferred a whole-genome phylogeny with maximum likelihood under a GTR model, with the Wuhan-Hu-1 sequence (NCBI RefSeq accession no. NC_045512.2) as the outgroup, using IQ-TREE v2.1.2 COVID-edition [34]. This approach was applied to haplotype IV (see Table 1), resulting in a reduced phylogeny with 30,314 tips.

Quantification of transmission of Alpha and Delta mixtures was conducted through an exhaustive breadth-first search, selecting clusters with at least 2 members and at least 90% target samples, utilizing the implementation by Ruiz-Rodriguez et al. [35]. The composition requirement was reduced from 100% to account for potential sequencing errors and ambiguous sample placements on the phylogenetic tree. Due to the increased dataset size and complexity, we quantified transmission of deletion repair Alpha and BA.1 viruses using a reimplementation of the algorithm that leverages

polytomies in a global-scale phylogeny for parallelization in an HPC environment, based on *phylobase* v0.8.10 (https://github.com/fmichonneau/phylobase). The complete pipeline is available as a Snakemake workflow at https://github.com/PathoGenOmics-Lab/transcluster (v1.0.0). The minimum number of independent emergences for each combinatorial haplotype was derived from the phylogenies by adding up the number of transmission clusters to the number of emergences that were not transmitted. We studied the location and time span of the transmission clusters as well (see Supplemental Figures 1 and 2). Host age comparisons between transmitted and non-transmitted samples included clusters with more than 2 members. We also conducted a detailed analysis of the age distribution of transmission clusters associated with combinatorial haplotype IV to exclude the influence of potential confounding factors (see Supplemental Figure 3).

To adequately compare the relative transmission success of the most prevalent haplotypes, we developed a method to estimate their transmission fitness for each cluster. This estimation involved calculating the ratio between the cluster's size and the number of sequences in GISAID that were collected between the first and last cases of the transmission cluster, with 7 days as padding at both sides of these windows to mitigate the effect of missing data and short cluster time spans. For clusters exhibiting cross-border transmission (involving samples from different geographic locations), we further divided the cluster time window into country-specific sub-windows. The denominator of the transmission fitness estimate was then calculated as the sum of the number of sequences in GISAID for each country-specific time window. This approach allowed us to account for variations in sampling efforts and prevalence across different time periods and geographic regions. Conducting the analyses without the 1-week padding or without differentiating country-specific sub-windows yielded significantly

different estimates, but the overall differences between haplotypes did not change (see Supplemental Figure 4). Differences in the distribution of the estimated transmission fitness between haplotypes were evaluated using Wilcoxon rank-sum tests. Statistical analyses were performed and visualized using R v4.1.2 [26] along with *tidyverse* v1.3.1 [27] and *ggpubr* v0.4.0 [36].

Biological characterization of BA.1 deletion repair

Combinations of deletion repairs of S:ΔH69/V70 and S:ΔV143/ΔY145 were introduced into a pCG1 plasmid encoding a codon-optimized BA.1 spike protein [37] by site-directed mutagenesis. All the constructs were verified by Sanger sequencing. Pseudotyped vesicular stomatitis virus (VSV) encoding a GFP reporter gene and carrying the different spike proteins was produced as previously reported [38]. To assess the effects on virus production, pseudotyped VSV carrying each construct were produced independently three times. The resulting viruses were then titrated by infecting Vero E6 cells (kindly provided by Dr. Luis Enjuanes; CNB, Spain) or Vero E6-TMPRSS2 cells (JCRB Cell bank catalogue code: JCRB1819) for 16 hours, followed by quantification of GFP-expressing infected cells using a live cell microscope (Incucyte SX5; Sartorius) to obtain the number of focus forming units (FFU) per millilitre. To assess thermal stability, 500 FFU of these pseudotyped viruses were incubated for 15 min at a range of temperature in a thermal cycler (30.4, 31.4, 33.0, 35.2, 38.2, 44.8, 47.0, 48.6 and 49.6°C; Biometra T one Gradient, Analytik Jena) and the surviving virus was used to infect VeroE6-TMPRSS2 cells for 16 h. The GFP signal in each well was then determined using a live-cell microscope (Incucyte SX5, Sartorius). The average GFP signal observed in mock-infected wells was subtracted from all infected wells, followed by standardization of the GFP signal to the average GFP signal from wells incubated at 30.4 °C. Finally, we fitted a three-parameter log-

logistic function to the data using the *drc* v3.0-1 R package (*LL.3* function) and calculated the temperature resulting in 50% reduction in virus infection (*ED* function). To assess the effects on neutralization by polyclonal sera, we used six sera from convalescent patients from the first COVID-19 wave in Spain and six sera from individuals that had been administered two doses of the BioNTech-Pfizer Comirnaty COVID-19 vaccine. The neutralization capacity of the sera was obtained as previously described on VeroE6-TMPRSS2 cells [37]. We used a previously described flow cytometry assay based on the use of polyclonal sera to examine surface expression [39]. Briefly, HEK293T cells were transfected with the different S mutants using the calcium chloride method. 24 h later, cells were detached using PBS with 1 mM EDTA, washed, and incubated on ice with different polyclonal sera (three from convalescent patients from the 1st COVID-19 wave in Spain and one from individuals that had been administered two doses of the BioNTech-Pfizer Comirnaty COVID-19 vaccine) at a 1:300 dilution in PBS containing 0.5 % BSA and 2 mM EDTA for 30 min. Next, cells were washed three times with PBS, stained with anti-IgG Alexa Fluor 647 (Thermo Fisher Scientific) at a 1:400 dilution, and analyzed by flow cytometry similarly treated un-transfected controls to set the threshold for positive cells. For cell-cell fusion assays we used a split Venus fluorescent protein system [40]. Briefly, HEK293T cells were grown overnight in 24 well plates (1.5 x 105 cells/well) using DMEM supplemented with 10 % FBS. After 24 hours, cells were transfected using Lipofectamine 2000 (Invitrogen) with 0.5 µg of either a 1:1 mixture of the S plasmids and a Jun-Nt Venus fragment (Addgene 22012) plasmid or a mixture of hACE2 plasmid (kindly provided by Dr. Markus Hoffman, German Primate Center, Goettingen/Germany) [41] and the Fos-Ct Venus fragment (Addgene 22013). After 24 h, cells were counted, and the S-transfected cells were mixed at a 1:1 ratio with ACE2-transfected cells and seeded in 96

well plates (3 x 104 cells/well) in 100 µL of media. Cells transfected with the Wuhan-Hu-1 served as a positive control, while cells transfected with hACE2 and Jun-Nt Venus were used as negative control. We obtained the GFP Integrated Intensity (GCU·µm²/image) in each condition using a live-cell imaging platform (Incucyte SX5, Sartorius) at 24 hours post mixing and standardized to the signal obtained from the positive control (Wuhan-Hu-1 spike protein). All experiments were performed at least three times in triplicates.

Statistical analyses were performed and visualized using R v4.1.2 [26] along with *tidyverse* v1.3.1 [27] and *ggpubr* v0.4.0 [36] to facilitate the analysis and enhance the visualization of the results. Comparisons were conducted utilizing t-tests (unpaired for all assays and paired for neutralization and surface expression, as we used the same sets of polyclonal sera in each experiment) after verifying that the data met the assumptions of normality using a Shapiro-Wilk test. To determine fold-change values, the ratio of group average values was calculated.

Results

Favoured and forbidden mixtures of NTD marker combinations in the Alpha and Delta variants

In the first part of this work, we focused on examining the differences in success between the Alpha and Delta variants in the NTD of the spike protein. This domain is characterized by recurrent deletions occurring in distinct independent lineages [8,14,15,17,42,43]. We identified 7,706 samples that carry a mixture of Alpha and Delta-defining mutations in this region, out of the 7.13 million sequences submitted to the GISAID's EpiCoV repository as of January 2022 (Supplemental Table 1). By comparing these two variants, we aimed to gain insights into potential variations in their

adaptive characteristics and overall performance in this specific region. In total, 14 combinatorial haplotypes were surveyed (termed I-XIV, Figure 1B). Differences in the number of observations were apparent, with the noticeable absence (n = 0) of haplotypes I, X, XII and XIV, and low prevalence (n < 100) of haplotypes II, III, VI, VIII, and IX. These haplotypes were omitted from further analysis, as they were not considered to be epidemiologically relevant. On the contrary, haplotypes IV, V, VII, XI, and XIII were continuously sampled throughout the time window of the variant takeover (Figure 1C and Figure 2A), with a remarkably high prevalence compared to the rest (Table 1). The haplotype network showed two well-separated haplotype groups around the two main variants of concern connected by the reference genome resembling a distance-based unrooted phylogeny with added alternative connections, i.e. mutational jumps (Figure 1C). Interestingly, three out of five of these combinatorial haplotypes (IV, V and XIII) bore repaired NTD deletions (Figure 1B).

Table 2. Description of six Alpha and Omicron BA.1 haplotypes with repaired NTD deletions in sites S:69/70 and S:144.

Haplotype	**Description**	**Observations**	**Transmission clusters**	**Minimum independent emergences**
Alpha RepΔ69/70	Alpha + repaired S:ΔH69/V70	722	0	722
Alpha RepΔ144	Alpha + repaired S:ΔY144	4,571	332	2,823
Alpha RepBoth	Alpha + repaired S:ΔH69/V70 & S:ΔY144	1,297	119	982
BA.1 RepΔ69/70	Omicron BA.1 + repaired S:ΔH69/V70	5,166	198	3,969
BA.1 RepΔ143/145	Omicron BA.1 + repaired S:ΔV143/Y145	637	27	537
BA.1 RepBoth	Omicron BA.1 + repaired S:ΔH69/V70 & S:ΔV143/Y145	737	63	453

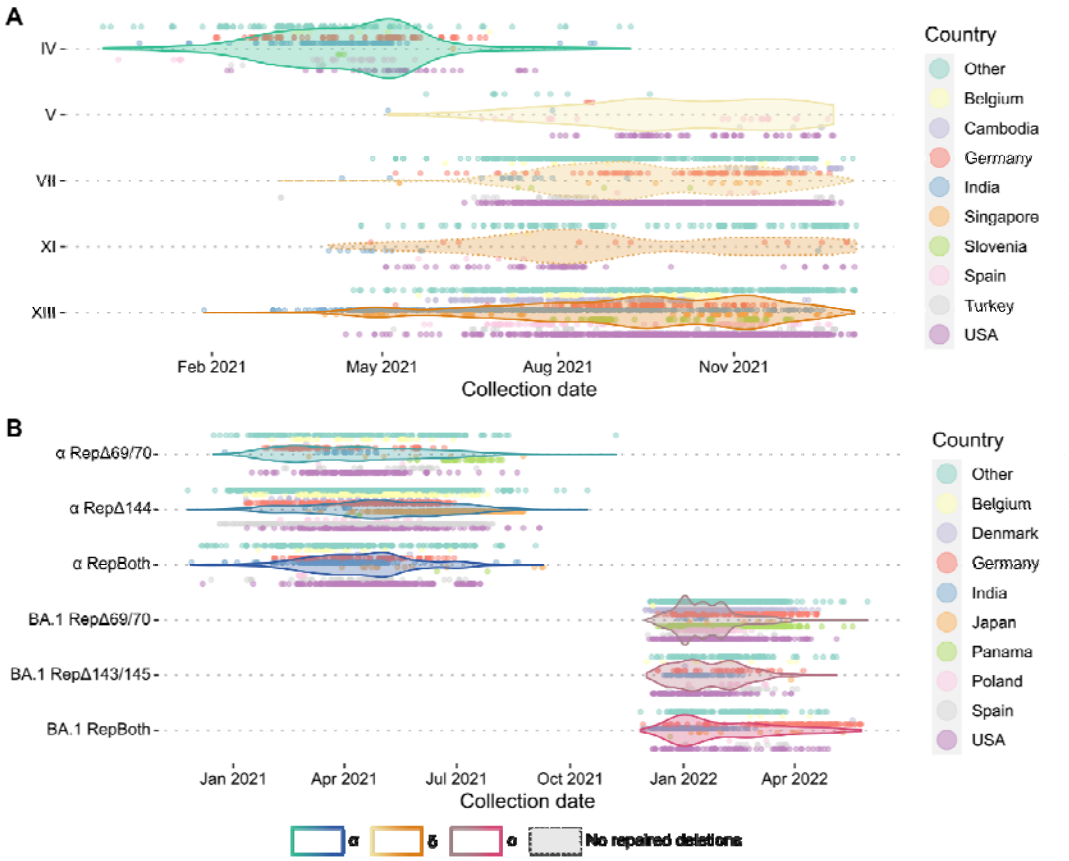

Figure 2. Sample collection timelines broken down by sampling location. The nine countries with the highest number of observations in each dataset are displayed for each panel. Each point represents a different sample assigned to the corresponding haplotype. (A) Observations of each Alpha and Delta NTD combinatorial haplotype with more than one hundred observations. (B) Observations of each deletion repair variant on Alpha and BA.1 (Omicron) background (see Table 2).

Nevertheless, estimates of prevalence can be distorted by several factors, including the geographical location and temporal specificity associated with both sequencing efforts and the distribution of lineages throughout the pandemic. Transmission, on the other hand, is usually considered as a proxy of viral fitness, because it is related to its basic reproductive number, reflecting the ability of the virus to replicate, persist and spread within hosts and in the population [44–47]. Therefore, we estimated the transmission of each haplotype by interrogating worldwide SARS-CoV-2 sequence data. This enabled the group-wise quantification of the minimum number of emergence events. We observed limited cross-border transmission, as 89 % of all transmission clusters were contained in a single country. The average within-cluster collection date window was 30 ± 44 days (Supplemental Figure 1). We found

vast differences in the number of clusters among haplotypes (see Table 1). In fact, we observed that this indicator, as well as the number of emergences, tended to increase linearly with the number of observations (both $R^2 = 1.0$, with $p = 2.2 \cdot 10^{-16}$ and $p = 4.7 \cdot 10^{-16}$, respectively). To mitigate this effect and better evaluate differences in transmission of the most frequent haplotypes, we estimated their transmission fitness by adjusting their cluster sizes to account for variation in sampling effort and prevalence across different time periods and geographic regions. We found that the two haplotypes with the highest median estimated transmission fitness bore repaired deletions, both on Delta (haplotype XIII) and Alpha (haplotype IV) genomic backgrounds (Figure 3A).

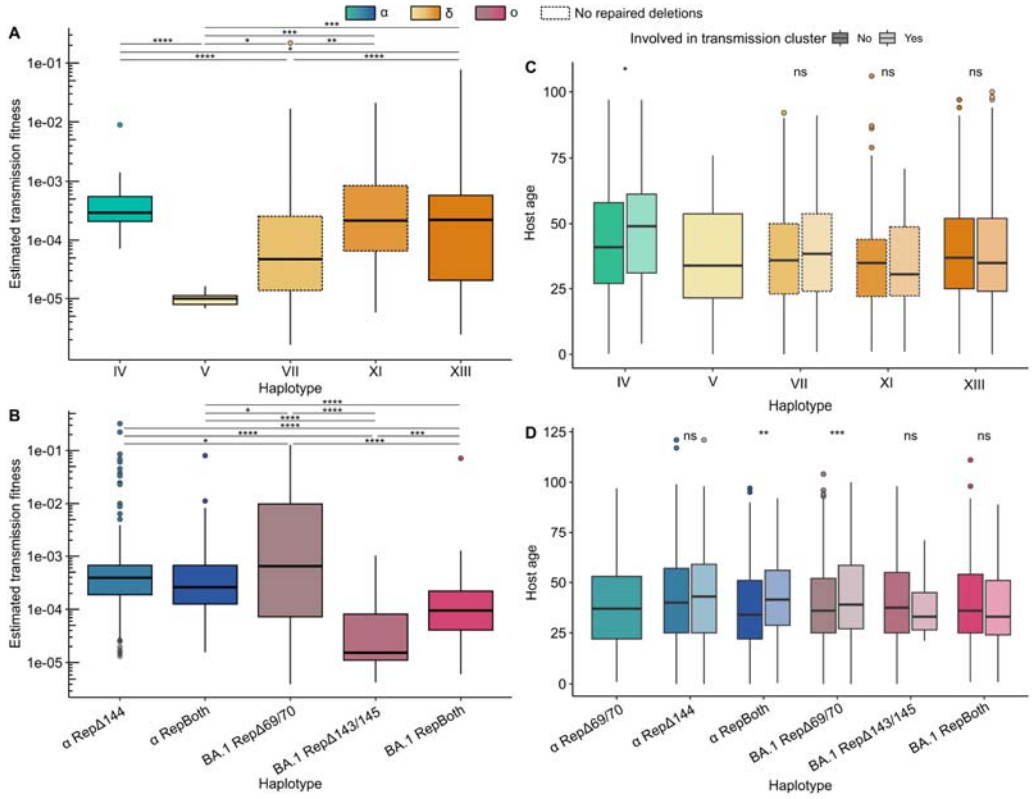

Figure 3. Differences in the estimated transmission fitness (A, B) and host age (C, D) between five Alpha and Delta combinatorial haplotype groups (A, C), and all six Alpha and Omicron BA.1 deletion repair haplotypes (B, D). Colour gradients indicate the variant representing each haplotype background. The differences between and within haplotypes were assessed using Wilcoxon rank-sum tests (ns: $p > 0.05$; *: $p \leq 0.05$; **: $p \leq 0.01$; ***: $p \leq 0.001$; ****: $p \leq 0.0001$). Only significant differences in panels A and B are displayed for clarity. Note that the haplotype corresponding to the Alpha variant with repaired S:ΔH69/V70 is missing from panel B due to its lack of transmission.

We sought to identify possible adaptive drivers among the clinical variables associated with these samples. We did not detect any significant differences related to

host sex. However, we found an association of host age with deletion repair in Alpha background: viruses assigned to haplotype IV infected older hosts (average 43 ± 21 years old) compared to the rest of NTD combinations (average difference of 6 ± 1 years; all $p < 0.005$, Wilcoxon rank-sum test), except V ($p = 0.085$, Wilcoxon rank-sum test). These differences are presented in Supplemental Figure 3A. We then assessed whether our results could be biased by the epidemiology and demography of the Alpha and Delta variants, which led to distinct spread patterns among different population groups. We confirmed that host age in haplotype IV was also higher when compared to samples that were collected in the same time window but not assigned to haplotype IV, or any of the remaining combinatorial haplotypes (both $p = 1.6 \cdot 10^{-6}$, Wilcoxon rank-sum test; Supplemental Figure 3B). Thus, sampling bias is unlikely to be the primary driver of the observed age differences between combinatorial haplotypes. Additionally, the associated host age was higher in samples that were involved in transmission events (Figure 3B). This points to the non-essentiality of S:ΔH69/V70 and S:ΔY144 for the infection of older individuals —who are often immunocompromised— in the Alpha background. However, it could be argued that certain outbreaks could skew the age distribution —for instance, if several large outbreaks occurred in elderly care facilities. To control for this factor, we analysed the within-location cluster size distribution at a regional level and found that hosts were generally older, but there were no dominant transmission events (Supplemental Figure 3C and 3D). Based on these findings, we find no evidence to suggest that age effects were driven by a few specific outbreaks.

To summarize, haplotypes featuring repaired deletions in the NTD exhibit an increased number of observations and transmission capabilities among all combinatorial possibilities, except for the established variants of concern. Furthermore, our findings show a pronounced difference in distribution and fitness between combinatorial

variation of Alpha and Delta, emphasizing the significance of genetic context in the evaluation of genotype-phenotype relationships.

Common patterns of NTD deletion repair confer different degrees of viral success on Alpha and Omicron BA.1 backgrounds

The Omicron BA.1 lineage emerged after November 2021 and rapidly outcompeted the Delta variant. The BA.1 lineage is defined by deletions S:ΔH69/V70 and S:ΔV143/Y145, which map to Alpha-defining deletions (see Figure 1A) and are known to also confer adaptive advantages in Alpha viruses [14,42]. Building upon our prior findings about the simultaneous repair of these two deletions in the Alpha variant (haplotype IV), we interrogated the possibility of similar repairs occurring in the Omicron BA.1 genetic background by investigating different patterns of deletion repair of epidemiological significance and the drivers behind their emergence. We performed an analogous global survey followed by a phylogenetic estimation of viral success of individual and dual deletion repairs in Alpha and Omicron BA.1 backgrounds. Due to the increased number of BA.1-defining markers compared to previous hegemonic variants of concern (see Figure 1A), we based the survey solely on the presence or absence (i.e. presence of the ancestral state) of the specific deletions of interest (Figure 1D). We identified 13,130 samples that carried repaired NTD deletions in Alpha or Omicron BA.1 backgrounds as of June 2022 (Supplemental Table 2).

In terms of the number of observations, there was a clear disparity in repair patterns between Alpha and Omicron BA.1 samples. The most frequently observed group consisted of samples exhibiting the repaired S:ΔH69/V70 in BA.1 background, followed by S:ΔY144 repair in Alpha, while the remaining repair patterns were significantly less frequently observed (Table 2). The overall number of observations of Omicron BA.1 viruses with double repair was predictably lower than that of the

predominant parental lineage, as the sampling window was nearly two-thirds narrower compared to that of other groups (Figure 2), but similar to that of combinatorial haplotypes in our previous analysis. We measured an average within-cluster collection date window of 25 ± 31 days, with 87 % of clusters showing within-border transmission (Supplemental Figure 2). The BA.1 haplotype bearing the S:ΔH69/V70 single repair had the highest median transmission fitness, while Alpha with the S:ΔH69/V70 single repair was not transmitted at all (Figure 3C). Incidentally, repair of S:ΔH69/V70 had no significant impact on transmission fitness when co-occurring with repaired S:ΔY144 in Alpha background. In turn, BA.1 with both repairs had a lower median transmission fitness than Alpha with both repairs. This was also the case of BA.1 with the S:ΔV143/Y145 single repair compared with the analogous Alpha with the S:ΔY144 single repair.

We further investigated the potential role of host age as an adaptive determinant of NTD deletion repair in Alpha and BA.1 background. We observed a significant difference in host age for the Alpha haplotype with both repairs, with transmitted samples exhibiting an elevated age compared to non-transmitted samples. This finding paralleled our previous observations for haplotype IV, which is genetically analogous (see Figures 1A and 1D for reference). However, this pattern did not hold for the Omicron BA.1 haplotype with both deletion repairs. In contrast, the single repair of S:ΔH69/V70 in BA.1 did show such an association (Figure 3D). In short, we observed a lack of correspondence in the population effect of NTD deletion repairs in both backgrounds. This fact, combined with the well-described effect of deletions in Alpha background (reviewed later in the Discussion section), suggested that the genetic background in which these deletions emerge is likely to have a differential functional impact.

NTD deletion repair in Omicron BA.1 background has a non-accumulative effect on viral phenotype

To investigate whether certain viral characteristics exist that act as drivers of deletion repair in Omicron, we analysed the effect of deletion repair patterns in the Omicron BA.1 using pseudotyped VSV bearing BA.1 spike proteins with repaired S:ΔH69/V70, S:ΔV143/Y145, or both. Specifically, we examined the efficiency of virus production, thermal stability, surface expression, susceptibility to antibody neutralization, and fusogenicity. To assess virus production, VSV was pseudotyped with all spike constructs under identical conditions at the same time, and the amount of virus produced titrated on Vero E6 cells (Figure 4A).

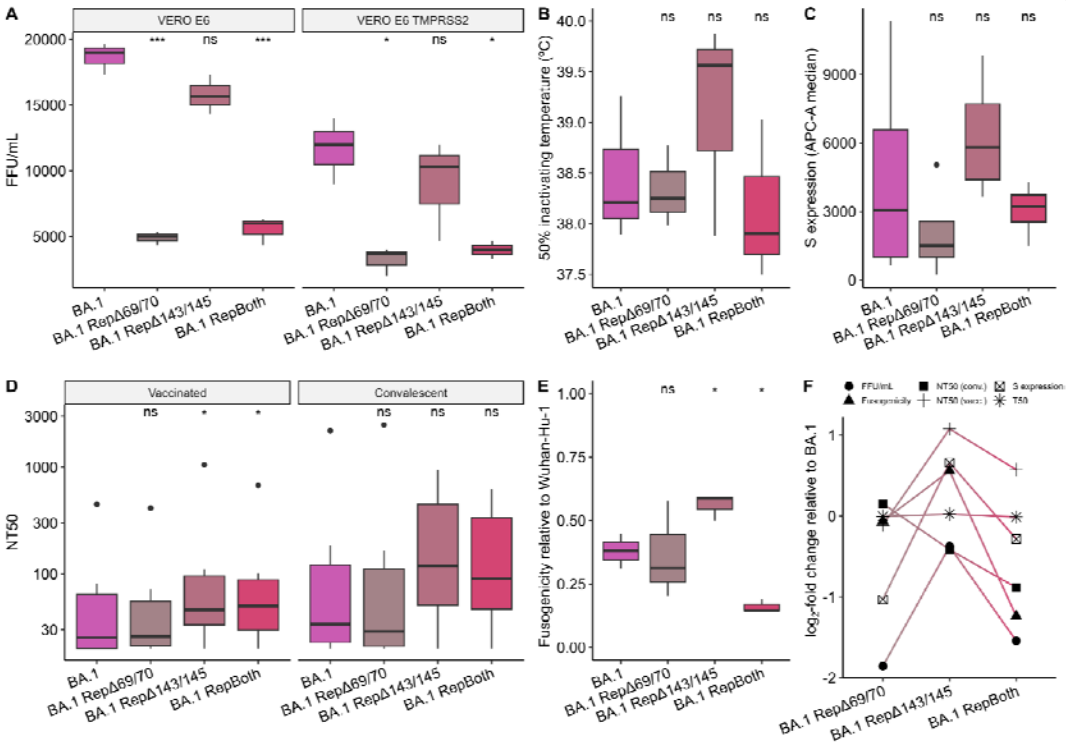

Figure 4. Experimental assessment of the effect of repairing NTD deletions in Omicron BA.1 background. (A) Comparison of viral titres obtained for each spike variant using pseudotyped VSV on the indicated cell line. (B) Comparison of the temperature resulting in 50% inactivation of pseudotyped VSV carrying the indicated S protein. (C) Surface expression of the spike protein in transfected HEK293T cells quantified using flow cytometry. (D) Reciprocal 50% neutralization titres (NT50) of sera from convalescent and vaccinated individuals. (E) Relative ability of each spike variant to drive cell-cell fusion relative to the Wuhan-Hu-1 spike. (F) Summary of the log2-transformed fold change for each spike variant relative to that of BA.1 in terms of virus production, fusogenicity, NT50, surface expression and 50% inactivating temperature. Subfigures A-E represent the median and interquartile range of at least three replicates. Differences were assessed using Wilcoxon rank-sum tests (ns: $p > 0.05$; *: $p \leq 0.05$; **: $p \leq 0.01$; ***: $p \leq 0.001$).

Virus production in Vero E6 cells was not affected by S:ΔV143/Y145 repair (0.85-fold; p = 0.06) but was significantly reduced upon repairing S:ΔH69/V70 (0.26-fold; $p = 6.6 \cdot 10^{-4}$) and the repair of both deletions (0.30-fold; $p = 1.6 \cdot 10^{-4}$). A similar effect was observed when the titre of the virus was evaluated in cells expressing the TMPRSS2 co-receptor, indicating a substantial negative effect of S:ΔH69/V70 repair by itself on virus production. We assessed whether this was the result of reduced thermal stability of the spike proteins by obtaining the temperature resulting in a 50% reduction of virus titre (i.e. 50% inactivation temperature; Figure 4B). No significant differences were observed, suggesting stability was not the driver of these differences.

Next, we examined whether the effects on virus production in Vero E6 cells stemmed from altered cell expression of the different constructs by flow cytometry, using polyclonal sera from four individuals (Figure 4C). We did not detect differences in the median cell surface expression of the spike protein between the canonical BA.1 protein and any deletion repair haplotype. However, the repair of S:ΔV143/Y145 resulted in higher expression than S:ΔH69/V70 (2.2-fold; p = 0.0030), and than the double repair (2.1-fold; p = 0.032), resembling the effect observed with virus production.

Then, we questioned whether deletion repair affected neutralization by polyclonal sera from convalescent donors from the first epidemic wave in Spain and those dually vaccinated with the Comirnaty mRNA vaccine (n = 6 each; Figure 4D). Significant increases in susceptibility to neutralization in sera from vaccinated individuals against viruses with repaired S:ΔV143/Y145 (2.1-fold; p = 0.011) or both S:ΔH69/V70 and S:ΔV143/Y145 repaired (1.5-fold; p = 0.020) were observed. A similar trend was observed with convalescent sera, but statistical significance was not reached (p = 0.19 and p = 0.42, respectively) due to higher intra sample variability. Thus, these results indicate that increased neutralization is driven by the repair of S:ΔV143/Y145 in BA.1

453 background.

454 Finally, as cell to cell spread via fusion of the plasma membrane could potentially
455 reduce exposure to neutralizing antibodies, we examined whether deletion repair could
456 alter the ability of the spike protein to fuse cells (Figure 4E). Interestingly, the repair of
457 S:ΔV143/Y145 increased cell fusion relative to the BA.1 spike protein (1.5-fold; p =
458 0.026), while the repair of both S:ΔH69/V70 and S:ΔV143/Y145 led to a decrease of
459 more than 50% in the average fusogenicity (0.42-fold; p = 0.021). The presence of
460 S:ΔH69/V70 by itself did not seem to play a role in cell-cell fusion in a BA.1
461 background.

462 **Discussion**

463 Deletions in the SARS-CoV-2 genome have a significant impact on viral adaptation and
464 fitness, often surpassing the effects of single nucleotide variants (SNVs)
465 [9,14,16,42,48]. In fact, deletions in the NTD region of the spike protein are fixed in
466 prominent viral variants. Thus, our understanding of the repair patterns of deletion is
467 crucial for elucidating the genetic and phenotypic characteristics of the variants of
468 concern. In this work, we demonstrate that repairing these deletions can alter viral
469 characteristics and potentially influence the success of specific viral haplotypes. Our
470 findings provide valuable insights into the genetic and phenotypic characteristics of
471 these variants, shedding light on the factors driving their emergence and transmission
472 dynamics.

473 We first examined the global distribution of SARS-CoV-2 samples carrying
474 different combinatorial patterns of spike NTD marker segments in the Alpha and Delta
475 variants. Upon examination of the genetic variations among the most prevalent
476 combinatorial haplotypes, it became apparent that only a single haplotype is assigned to

the Alpha variant: the one bearing repaired S:ΔH69/V70 and S:ΔY144 (IV, n = 736). The fact that only one combinatorial haplotype was derived from the Alpha lineage — compared to all the rest derived from Delta— could imply that Alpha viruses were more constrained regarding NTD mutation interactions, although the difference in prevalence and time frames between variants may have also played a role. Indeed, Alpha bearing S:ΔE156/F157-R158G (X) and Delta bearing S:ΔH69/V70 and S:ΔY144 (XI) share an NTD that contains all Alpha and Delta lineage-defining changes. However, the former has never been detected (n = 0), while the latter is the fourth most abundant combinatorial haplotype (n = 304). The fact that an Alpha spike with S:ΔE156/F157-R158G has never been recorded might be suggestive of an epistatic incompatibility with infection success. It might also be related to these sites mapping to the same b-hairpin in the NTD supersite as S:144 [49], which is typically deleted in Alpha. This also ties to previous suggestions of host and variant-specific structural restrictions being imposed strictly by the length of NTD loops [9] and emphasizes the significance of genetic context in genotype-phenotype relationships.

Furthermore, we investigated clinical variables associated with these samples, and found an association between host age and deletion repair in the Alpha background, with Alpha viruses with repaired S:ΔH69/V70 and S:ΔY144 being more prevalent among older individuals. This suggests a non-essential role of certain deletions in elderly hosts. The association between deletion repair and age was further supported by the higher age in samples involved in transmission events. Besides, the association was not driven by just a few outbreaks. Focusing on these deletions, both S:ΔH69/V70 and S:ΔY144 map to the NTD antigenic supersite [8,49] and have been recurrently identified in immunocompromised individuals with chronic infections before widespread vaccination against COVID-19 [7,10,11,14]. Earlier in the pandemic, Meng

et al. [42] reported the independent emergence of S:ΔH69/V70 after the occurrence of infectivity-impairing amino acid changes that, in turn, could promote immune escape or stronger receptor-binding affinity. Interestingly, this study also showed that S:ΔH69/V70 compensated for the deleterious effect of these mutations in a surrogate system by increasing viral infectivity and cell-cell fusogenicity in Alpha background. Previous research has pointed out that variation in the NTD can act as a fine-tuning vehicle for accommodating diverse pressures [9]. During the initial emergence of S:ΔH69/V70 itself, the lack of vaccination limited selective pressures [50]. In this perspective, repair of S:ΔH69/V70 and S:ΔY144 might be able to stay as a permissive mutation in immunocompromised patients with a minimally constrained adaptive landscape. This, in turn, could facilitate the further acquisition of otherwise deleterious mutations.

In the second part of our study, we focused on the emergence of repaired deletions in the Omicron BA.1 lineage compared to the Alpha variant, finding striking differences. We detected a higher number of cases with repaired deletions in lineage BA.1 than in Alpha. Interestingly, BA.1 with repaired S:ΔH69/V70 exhibited the highest transmission fitness, while Alpha with the same repair did not transmit to any extent. Additionally, we did not detect any age association with repaired deletion in Omicron background like we detected in Alpha. This could be due to several factors, such as the influence of viral genetic context or the different immune status of the population when Alpha and Omicron BA.1 were circulating.

Survey efforts are expected to be biased by geographic and temporal differences in sequencing efforts and lineage prevalence, and thus the phylogenetic analysis of transmission and emergence may not capture all transmission events or account for other epidemiological factors that affect viral success. However, phylogenetic

estimation of transmission and emergence rates effectively enabled us to overcome these biases and better understand viral spread dynamics. The consistency between our two methodologies for estimating transmission (operating under maximum likelihood and maximum parsimony) reassures our inferences about the genetic relationships and evolutionary dynamics within our dataset. In summary, our exhaustive approach revealed a wide variation in the number of observations of these haplotypes, highlighting their differences in their success and persistence over time. Overall, these results demonstrate the distinct effect on viral success of common NTD markers depending on their genomic background and lineage, highlighting once more the critical importance of the genetic context when describing the genotype-phenotype relationship of mutations.

To gain further insights into the viral characteristics driving deletion repair in Omicron BA.1 background, we conducted an array of *in vitro* experiments using spike proteins bearing different deletion repair patterns. We assessed virus production, thermal stability, surface expression, susceptibility to neutralization, and fusogenicity of each spike haplotype. The similarity in cell surface expression of the engineered spike protein compared to the canonical spike rejected the possibility of other viral phenotypes in our surrogate system being dependent solely on this factor. Our results show that deletion repair patterns have an impact on virus production and infectivity, with repair of S:ΔH69/V70 leading to reduced virus titres by itself. This viral phenotype parallels that of the Alpha spike protein, with earlier studies reporting that repair of S:ΔH69/V70 in this background results in a decrease in infectivity and spike incorporation into virions [42]. We did not observe differences in thermal stability with the BA.1 protein, suggesting that the effect of S:ΔH69/V70 and S:ΔV143/Y145 repair does not affect full-protein stability.

552 However, deletion repair did affect sera neutralization. In our BA.1 background,
553 repaired S:ΔV143/Y145 resulted in an increase in sensitivity to neutralization by
554 polyclonal sera obtained from vaccinated individuals, in line with previous findings of
555 S:ΔY144 facilitating escape from NTD neutralizing antibodies in the Alpha background
556 [51]. The same was not observed for sera from convalescent, "first-wave" patients. In
557 turn, repair of S:ΔH69/V70 did not affect neutralization. This paralleled prior research
558 on Alpha and other earlier variants showing that neither S:ΔH69/V70 nor its repair
559 influenced the spike sensitivity to NTD neutralizing antibodies [9,42,51]. These results
560 point to the variant-independent influence of NTD variability on immune effects.
561 Incidentally, the association of higher host age with viral transmission of S:ΔH69/V70
562 repair in BA.1 background may be attributed to a lack immune selection at the S:69/70
563 sites following vaccination efforts. This is coherent with deletion repair being able to
564 emerge in elderly patients with reduced adaptive constraints, potentially facilitating
565 functional adaptation during prolonged infections.
566 Research has shown that cell-cell fusion and subsequent formation of syncytia
567 can be mediated by viral membrane glycoproteins [52,53] such as the spike protein
568 found in SARS-CoV-2. In fact, this is a key feature of SARS-CoV-2 infection
569 [41,54,55]. In line with prior research, our results point to a lower fusogenicity of the
570 unaltered BA.1 spike protein compared with Wuhan-Hu-1, which has been attributed to
571 its inefficient spike cleavage and unfavoured TMPRSS2-mediated cell entry [42].
572 Interestingly, we found that deletion repair patterns exerted a non-accumulative
573 influence on the fusogenicity of the BA.1 spike protein. Proteins with repaired
574 S:ΔV143/Y145 alone promoted higher fusogenicity, while S:ΔH69/V70 alone did not
575 have a significant impact. However, repair of both deletions led to a significant decrease
576 in fusogenicity. In contrast, the spike protein of the Alpha variant has been shown to

exhibit an increased potential for cell-to-cell fusion. Nevertheless, upon S:ΔH69/V70 repair, the fusogenicity is reduced to levels comparable to the Wuhan-Hu-1 spike [42].

Regardless, our surrogate system might not fully reflect the complex interactions of the spike protein with host cells and the immune system *in vivo*, and other regions of the spike protein or the viral genome might also contribute to SARS-CoV-2 fitness and adaptation. Besides, despite our attempts to account for demographic differences between variants, our analysis could be affected by the non-random nature of the sequences available in GISAID. The GISAID SARS-CoV-2 database is a remarkable initiative and tool for monitoring, which has greatly advanced the knowledge of the pandemic. Still, it also has significant limitations in terms of the origin and distribution of the sequences that are submitted, and it cannot be assumed to reflect unbiased genome surveillance. We also acknowledge that our suggestion of epistatic incompatibility and genetic constraints is not conclusive, since we have limited our analysis to a specific region of the S gene to narrow the number of possible combinatorial haplotypes. A more comprehensive analysis could consider the whole gene, or even the entire genome, thus requiring more computational and experimental resources and the availability of sufficiently curated sequence data.

Our study highlights the importance of deletion repair in the spike protein NTD in SARS-CoV-2 and its impact on viral fitness, transmission, and clinical characteristics depending on the genetic context. We have provided novel insights into NTD marker combinations in the Alpha and Delta variants, the repair patterns of NTD deletions in the Alpha and Omicron BA.1 backgrounds, and the phenotypic consequences of NTD deletion repair in the Omicron BA.1 background. There have been efforts to characterize the variability of the S protein NTD through full swap assays, comparing the ancestral (Wuhan-Hu-1) spike with that of Alpha and Omicron BA.1 [9], Delta and

Omicron BA.1 and BA.2 [42] and even with more distant virus like SARS-CoV [56]. However, to our best knowledge, no previous study has performed an exhaustive combinatorial approach to characterize each separate marker in each genomic context. By comparing the outcomes of deletion repair in different variants, we aim to gain insights into the specific adaptive characteristics and genomic context that influence the genotype-phenotype relationships. Our findings reveal that repair of specific deletions in different genetic backgrounds can be driven by distinct phenotypic traits, such as enhanced viral transmission, altered host age distribution, and changes in viral characteristics, including fusogenicity, infectivity and susceptibility to neutralization. Understanding these genotype-phenotype relationships can provide valuable insights into the evolutionary dynamics and adaptation of SARS-CoV-2 variants, aiding in the development of effective control strategies and therapeutic interventions. Future studies building upon these findings will contribute to the development of effective strategies to monitor and mitigate the impact of NTD deletions in emerging SARS-CoV-2 variants.

Acknowledgements

We thank Dr. Óscar González Recio (INIA-CSIC, Spain) for providing sequencing data from a representative BA.1 sample with S:142G and repaired NTD deletions (GISAID accession code: EPI_ISL_9805648), Dr. Luis Enjuanes (CNB, Spain) for providing Vero E6 cells, and Dr. Markus Hoffman (German Primate Center, Goettingen/Germany) for providing the hACE2 plasmid. We also thank Francisco José Martínez Martínez (IBV-CSIC, Spain) for comments on phylogeny visualization and analysis of transmission clusters. The computations were performed on the HPC cluster Garnatxa at the Institute for Integrative Systems Biology (I^2SysBio). The I^2SysBio is a is a joint collaborative research institute involving the University of Valencia (UV) and the Spanish National Research Council (CSIC). We gratefully acknowledge all data

627 contributors, i.e., the Authors and their Originating laboratories responsible for
628 obtaining the specimens, and their Submitting laboratories for generating the genetic
629 sequence and metadata and sharing via the GISAID Initiative, on which this research is
630 based.

631 **Funding details**
632 This research work was funded by the European Commission – NextGenerationEU
633 (Regulation EU 2020/2094), through CSIC's Global Health Platform (PTI+ Salud
634 Global) to MC, RG, IC and FGC. MAH is supported by the Generalitat Valenciana and
635 the European Social Fund “ESF Investing in your future” through grant
636 CIACIF/2022/333. This work was also a part of projects CNS2022-135116 (MC) and
637 CNS2022-135100 (RG) funded by MCIN/AEI/10.13039/501100011033 and the
638 European Union NextGenerationEU/PRTR.

639 **Ethics statement**
640 Sera samples for the biological characterization of BA.1 deletion repair were obtained
641 from the La Fe University and Polytechnic Hospital of Valencia and were collected
642 after informed written consent had been obtained, with approval by the ethical
643 committee and institutional review board (registration number 2020-123-1).

644 **Contributions**
645 MAH, BND, PRR and MC conceived the theoretical framework. MAH and BND
646 devised the initial idea. MAH and MC conceived the second part of the study. MAH
647 implemented and performed the data retrieval, data curation and computations. JZ, BG,
648 and RG designed the experiments. JZ and BG performed the experiments. MAH carried
649 out the statistical analyses. MG and CAG contributed biological samples. MAB did
650 project management. MAH drafted the manuscript with support from BDN, PRR, RG

and MC. FGC aided in interpreting the results. MAH, PRR, BND, RG, FGC, IC and MC discussed the results and commented on the manuscript. MC supervised the project.

Declaration of interest statement

The authors report there are no competing interests to declare.

References

[1] Rambaut A, Holmes EC, O'Toole Á, et al. A dynamic nomenclature proposal for SARS-CoV-2 lineages to assist genomic epidemiology. Nat Microbiol. 2020;5:1403–1407.

[2] Chen Z, Chong KC, Wong MCS, et al. A global analysis of replacement of genetic variants of SARS-CoV-2 in association with containment capacity and changes in disease severity. Clin Microbiol Infect. 2021;27:750–757.

[3] da Silva Francisco Junior R, Lamarca AP, de Almeida LGP, et al. Turnover of SARS-CoV-2 Lineages Shaped the Pandemic and Enabled the Emergence of New Variants in the State of Rio de Janeiro, Brazil. Viruses. 2021;13:2013.

[4] Letko M, Marzi A, Munster V. Functional assessment of cell entry and receptor usage for SARS-CoV-2 and other lineage B betacoronaviruses. Nat Microbiol. 2020;5:562–569.

[5] Harvey WT, Carabelli AM, Jackson B, et al. SARS-CoV-2 variants, spike mutations and immune escape. Nat Rev Microbiol. 2021;19:409–424.

[6] V'kovski P, Kratzel A, Steiner S, et al. Coronavirus biology and replication: implications for SARS-CoV-2. Nat Rev Microbiol. 2021;19:155–170.

[7] Kemp SA, Collier DA, Datir RP, et al. SARS-CoV-2 evolution during treatment of chronic infection. Nature. 2021;592:277–282.

[8] McCallum M, De Marco A, Lempp FA, et al. N-terminal domain antigenic mapping reveals a site of vulnerability for SARS-CoV-2. Cell. 2021;184:2332-2347.e16.

[9] Cantoni D, Murray MJ, Kalemera MD, et al. Evolutionary remodelling of N-terminal domain loops fine-tunes SARS-CoV-2 spike. EMBO Rep. 2022;n/a:e54322.

[10] Avanzato VA, Matson MJ, Seifert SN, et al. Case Study: Prolonged Infectious SARS-CoV-2 Shedding from an Asymptomatic Immunocompromised Individual with Cancer. Cell. 2020;183:1901-1912.e9.

[11] Choi B, Choudhary MC, Regan J, et al. Persistence and Evolution of SARS-CoV-2 in an Immunocompromised Host. N Engl J Med. 2020;383:2291–2293.

[12] Minskaia E, Hertzig T, Gorbalenya AE, et al. Discovery of an RNA virus 3'->5' exoribonuclease that is critically involved in coronavirus RNA synthesis. Proc Natl Acad Sci U S A. 2006;103:5108–5113.

[13] Denison MR, Graham RL, Donaldson EF, et al. Coronaviruses: an RNA proofreading machine regulates replication fidelity and diversity. RNA Biol. 2011;8:270–279.

[14] McCarthy KR, Rennick LJ, Nambulli S, et al. Recurrent deletions in the SARS-CoV-2 spike glycoprotein drive antibody escape. Science. 2021;371:1139–1142.

[15] Planas D, Veyer D, Baidaliuk A, et al. Reduced sensitivity of SARS-CoV-2 variant Delta to antibody neutralization. Nature. 2021;596:276–280.

[16] Mishra T, Dalavi R, Joshi G, et al. SARS-CoV-2 spike E156G/Δ157-158 mutations contribute to increased infectivity and immune escape. Life Sci Alliance. 2022;5:e202201415.

[17] Tamura T, Ito J, Uriu K, et al. Virological characteristics of the SARS-CoV-2 XBB variant derived from recombination of two Omicron subvariants. Nat Commun. 2023;14:2800.

[18] Zhang X, Chen L-L, Ip JD, et al. Omicron sublineage recombinant XBB evades neutralising antibodies in recipients of BNT162b2 or CoronaVac vaccines. Lancet Microbe. 2022;0.

[19] Mykytyn AZ, Rosu ME, Kok A, et al. Antigenic mapping of emerging SARS-CoV-2 omicron variants BM.1.1.1, BQ.1.1, and XBB.1. Lancet Microbe. 2023;0.

[20] Uriu K, Ito J, Zahradnik J, et al. Enhanced transmissibility, infectivity, and immune resistance of the SARS-CoV-2 omicron XBB.1.5 variant. Lancet Infect Dis. 2023;23:280–281.

[21] Yue C, Song W, Wang L, et al. ACE2 binding and antibody evasion in enhanced transmissibility of XBB.1.5. Lancet Infect Dis. 2023;23:278–280.

[22] Khare S, Gurry C, Freitas L, et al. GISAID's Role in Pandemic Response. China CDC Wkly. 2021;3:1049–1051.

[23] Hadfield J, Megill C, Bell SM, et al. Nextstrain: real-time tracking of pathogen evolution. Bioinformatics. 2018;34:4121–4123.

[24] Aksamentov I, Roemer C, Hodcroft EB, et al. Nextclade: clade assignment, mutation calling and quality control for viral genomes. J Open Source Softw. 2021;6:3773.

[25] Sanderson T, Barrett JC. Variation at Spike position 142 in SARS-CoV-2 Delta genomes is a technical artifact caused by dropout of a sequencing amplicon. Wellcome Open Res. 2021;6:305.

[26] R Core Team. R: A Language and Environment for Statistical Computing [Internet]. Vienna, Austria: R Foundation for Statistical Computing; 2021. Available from: https://www.R-project.org/.

[27] Wickham H, Averick M, Bryan J, et al. Welcome to the tidyverse. J Open Source Softw. 2019;4:1686.

[28] Ou J, Zhu LJ. trackViewer: a Bioconductor package for interactive and integrative visualization of multi-omics data. Nat Methods. 2019;16:453–454.

[29] Paradis E. Analysis of haplotype networks: The randomized minimum spanning tree method. Methods Ecol Evol. 2018;9:1308–1317.

[30] Paradis E. pegas: an R package for population genetics with an integrated–modular approach. Bioinformatics. 2010;26:419–420.

[31] McBroome J, Thornlow B, Hinrichs AS, et al. A Daily-Updated Database and Tools for Comprehensive SARS-CoV-2 Mutation-Annotated Trees. Mol Biol Evol. 2021;38:5819–5824.

[32] Turakhia Y, Thornlow B, Hinrichs AS, et al. Ultrafast Sample placement on Existing tRees (UShER) enables real-time phylogenetics for the SARS-CoV-2 pandemic. Nat Genet. 2021;53:809–816.

[33] Fu L, Niu B, Zhu Z, et al. CD-HIT: accelerated for clustering the next-generation sequencing data. Bioinformatics. 2012;28:3150–3152.

[34] Minh BQ, Schmidt HA, Chernomor O, et al. IQ-TREE 2: New Models and Efficient Methods for Phylogenetic Inference in the Genomic Era. Teeling E, editor. Mol Biol Evol. 2020;37:1530–1534.

[35] Ruiz-Rodriguez P, Francés-Gómez C, Chiner-Oms Á, et al. Evolutionary and Phenotypic Characterization of Two Spike Mutations in European Lineage 20E of SARS-CoV-2. mBio. 2021;12:e02315-21.

[36] Kassambara A. ggpubr: “ggplot2” Based Publication Ready Plots. 2023. Available from: https://rpkgs.datanovia.com/ggpubr/.

[37] Giménez E, Albert E, Zulaica J, et al. Severe Acute Respiratory Syndrome Coronavirus 2 Adaptive Immunity in Nursing Home Residents Following a Third Dose of the Comirnaty Coronavirus Disease 2019 Vaccine. Clin Infect Dis. 2022;75:e865–e868.

[38] Gozalbo-Rovira R, Gimenez E, Latorre V, et al. SARS-CoV-2 antibodies, serum inflammatory biomarkers and clinical severity of hospitalized COVID-19 patients. J Clin Virol. 2020;131:104611.

[39] Grzelak L, Temmam S, Planchais C, et al. A comparison of four serological assays for detecting anti–SARS-CoV-2 antibodies in human serum samples from different populations. Sci Transl Med. 2020;12:eabc3103.

[40] García-Murria MJ, Expósito-Domínguez N, Duart G, et al. A Bimolecular Multicellular Complementation System for the Detection of Syncytium Formation: A New Methodology for the Identification of Nipah Virus Entry Inhibitors. Viruses. 2019;11:229.

[41] Hoffmann M, Kleine-Weber H, Schroeder S, et al. SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. Cell. 2020;181:271-280.e8.

[42] Meng B, Kemp SA, Papa G, et al. Recurrent emergence of SARS-CoV-2 spike deletion H69/V70 and its role in the Alpha variant B.1.1.7. Cell Rep. 2021;35:109292.

[43] McCallum M, Walls AC, Sprouse KR, et al. Molecular basis of immune evasion by the Delta and Kappa SARS-CoV-2 variants. Science. 2021;374:1621–1626.

[44] Domingo E. Chapter 7 - Long-Term Virus Evolution in Nature. In: Domingo E, editor. Virus Popul. Boston: Academic Press; 2016. p. 227–262.

[45] Achaiah NC, Subbarajasetty SB, Shetty RM. R0 and Re of COVID-19: Can We Predict When the Pandemic Outbreak will be Contained? Indian J Crit Care Med Peer-Rev Off Publ Indian Soc Crit Care Med. 2020;24:1125–1127.

[46] Zhao L, Wymant C, Blanquart F, et al. Phylogenetic estimation of the viral fitness landscape of HIV-1 set-point viral load. Virus Evol. 2022;8:veac022.

[47] Chaudhry MRA. Chapter 5 - Coronavirus infection outbreak: comparison with other viral infection outbreak. In: Qureshi AI, Saeed O, Syed U, editors. Coronavirus Dis. Academic Press; 2022. p. 47–57.

[48] Venkatakrishnan AJ, Anand P, Lenehan PJ, et al. Expanding repertoire of SARS-CoV-2 deletion mutations contributes to evolution of highly transmissible variants. Sci Rep. 2023;13:257.

[49] Cerutti G, Guo Y, Zhou T, et al. Potent SARS-CoV-2 neutralizing antibodies directed against spike N-terminal domain target a single supersite. Cell Host Microbe. 2021;29:819-833.e7.

[50] MacLean OA, Lytras S, Weaver S, et al. Natural selection in the evolution of SARS-CoV-2 in bats created a generalist virus and highly capable human pathogen. PLoS Biol. 2021;19:e3001115.

[51] Graham C, Seow J, Huettner I, et al. Neutralization potency of monoclonal antibodies recognizing dominant and subdominant epitopes on SARS-CoV-2 Spike is impacted by the B.1.1.7 variant. Immunity. 2021;54:1276-1289.e6.

[52] Li W, Moore MJ, Vasilieva N, et al. Angiotensin-converting enzyme 2 is a functional receptor for the SARS coronavirus. Nature. 2003;426:450–454.

[53] Tseng C-TK, Tseng J, Perrone L, et al. Apical Entry and Release of Severe Acute Respiratory Syndrome-Associated Coronavirus in Polarized Calu-3 Lung Epithelial Cells. J Virol. 2005;79:9470–9479.

[54] Buchrieser J, Dufloo J, Hubert M, et al. Syncytia formation by SARS-CoV-2-infected cells. EMBO J. 2020;39:e106267.

[55] Bussani R, Schneider E, Zentilin L, et al. Persistence of viral RNA, pneumocyte syncytia and thrombosis are hallmarks of advanced COVID-19 pathology. EBioMedicine. 2020;61:103104.

[56] Qing E, Kicmal T, Kumar B, et al. Dynamics of SARS-CoV-2 Spike Proteins in Cell Entry: Control Elements in the Amino-Terminal Domains. mBio. 2021;12:10.1128/mbio.01590-21.