Perspectives: SARS-CoV-2 Spike Convergent Evolution as a Guide to Explore Adaptive Advantage

Jiri Zahradník[1], Jaroslav Nunvar[2,3] and Gideon Schreiber[1]*

[1] Department of Biomolecular Sciences, Weizmann Institute of Science, Rehovot, Israel, [2] Department of Genetics and Microbiology, Faculty of Science, Charles University, Prague, Czechia, [3] Biotechnology and Biomedicine Centre of the Academy of Sciences and Charles University (BIOCEV), Vestec, Czechia

Viruses rapidly co-evolve with their hosts. The 9 million sequenced SARS-CoV-2 genomes by March 2022 provide a detailed account of viral evolution, showing that all amino acids have been mutated many times. However, only a few became prominent in the viral population. Here, we investigated the emergence of the same mutations in unrelated parallel lineages and the extent of such convergent evolution on the molecular level in the spike (S) protein. We found that during the first phase of the pandemic (until mid 2021, before mass vaccination) 31 mutations evolved independently ≥3-times within separated lineages. These included all the key mutations in SARS-CoV-2 variants of concern (VOC) at that time, indicating their fundamental adaptive advantage. The omicron added many more mutations not frequently seen before, which can be attributed to the synergistic nature of these mutations, which is more difficult to evolve. The great majority (24/31) of S-protein mutations under convergent evolution tightly cluster in three functional domains; N-terminal domain, receptor-binding domain, and Furin cleavage site. Furthermore, among the S-protein receptor-binding motif mutations, ACE2 affinity-improving substitutions are favoured. Next, we determined the mutation space in the S protein that has been covered by SARS-CoV-2. We found that all amino acids that are reachable by single nucleotide changes have been probed multiple times in early 2021. The substitutions requiring two nucleotide changes have recently (late 2021) gained momentum and their numbers are increasing rapidly. These provide a large mutation landscape for SARS-CoV-2 future evolution, on which research should focus now.

Keywords: SARS-CoV-2, convergent evolution, mutations, virus, spike (S) protein

RESULTS AND DISCUSSION

Early Stages of SARS-CoV-2 Evolution

In the initial stages of the SARS-CoV-2 pandemic, virus evolution was shaped by selection imposed by a naïve host and the environment, resulting in new variants with adaptive advantage rapidly taking over previous strains. At this stage, global attention was focused on a number of major variants of concern: alpha (B.1.1.7), initially prominent in the United Kingdom, beta (B.1.351), discovered in South Africa, gamma (P.1) which has spread rapidly in the State of Amazonas and

delta (B.1.617.2) spreading from India. With the widespread introduction of vaccines during the first half of 2021, SARS-CoV-2 evolution shifted towards immune evading variants, most notable omicron BA1 discovered in South Africa and most recently omicron BA2. Independent acquisitions of S-protein substitutions L452R, E484K/Q, N501Y, and Q677H in these and other lineages were analyzed in great detail (Cherian et al., 2021; Hodcroft et al., 2021; Lemmermann et al., 2021; Martin et al., 2021; Zhou et al., 2021). The intra-host SARS-CoV-2 genomic diversity (Ramazzotti et al., 2021) and the viral evolution in immunocompromised patients (Choi et al., 2020; Clark et al., 2021; Tarhini et al., 2021; Truong et al., 2021) also received a lot of attention. Yet studies focused on convergent (syn. parallel) evolution highlighting new potential mutations of interest and their combination are scarce. Convergent evolution is evidence that natural selection plays a pivotal role in the emergence of remarkable patterns of independently evolving mutations that are arising independently within the members of different lineages (Martin et al., 2021). Here, we analyzed the convergent evolution of SARS-CoV-2 spike protein (S-protein) amino acid (AA) changes which have emerged independently since early 2020 in three or more prominent lineages. In addition, an exhaustive analysis of all possible S-protein receptor-binding motif substitutions, which are reachable by single- and double-nucleotide mutations, was conducted with respect to their global presence and binding effect. Our findings map the peculiarities of the SARS-CoV-2 mutational landscape and reinforce the need of careful monitoring of SARS-CoV-2 evolution.

Mutations With Putative Fitness Advantage Accumulate Independently in Multiple SARS-CoV-2 Lineages

In the GISAID database (Shu and McCauley, 2017), only five percent of the S-protein amino acids (AAs) show mutations in more than 100,000 genomic sequences (out of 9 mil. genomic sequences in GISAID; March 15, 2022). By theory, the fitness advantage of a mutation translates into the increased representation of the viral lineage carrying it; the extreme example being a selective sweep/lineage replacement. This happened with the D614G mutation, which quickly became dominant during the early stages of the pandemic and which positive impact on the virus fitness is well recognized (Volz et al., 2021). However, most of the conspicuous mutations which keep rising in prevalence have emerged by convergent evolution and many of them are present in dominant delta and subsequently omicron lineages. The fact that the same mutations emerge independently in different viral lineages is a standalone, strong indicator that these changes confer an adaptive advantage towards the virus infectivity and easier spread in the population. Similarly, a fixation of multiple different amino acids with chemical similarities implicates the ongoing optimization at the given site. **Figure 1** shows the distribution of convergent S-protein AA mutations which emerged at least three times independently during the period of a dramatic SARS-CoV-2 evolution between autumn 2020 and spring 2021. This comparison identified a total of 31 S-protein sites under convergent evolution. As expected, the most concerning SARS-CoV-2 lineages carry the heaviest burden of convergent S-
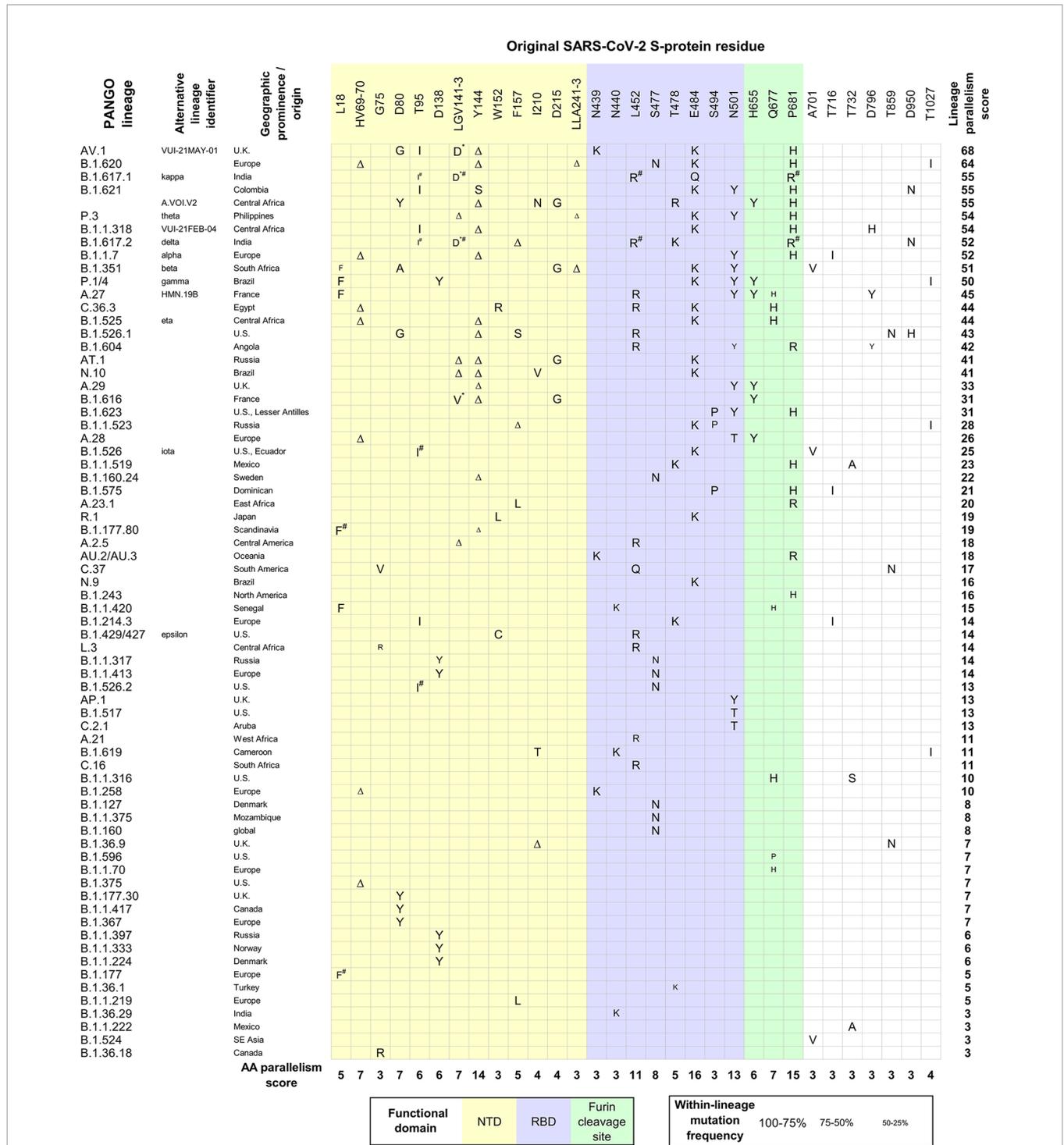
protein mutations (**Figure 1**). Since this analysis (purposefully) neglects mutations in other viral genes and their putative effects, it demonstrates that the S-protein mutations *per se* are strong determinants of the success of viral lineages hosting them, although rapid convergent evolution can be seen in other genes as well (De Maio et al., 2021; Majumdar and Niyogi, 2021).

Convergently Evolved Mutations Cluster in Specific Locations in the Spike Protein

Mapping the convergently evolved mutations, at the time of the highest SARS-CoV-2 strains variability (between autumn 2020 and spring 2021), on the S-protein structure shows that the AAs under the strongest parallelism concentrate in three hotspots: N-terminal domain (NTD), receptor-binding domain (RBD) and the Furin cleavage site (**Figure 2A**). Such conspicuous nonrandom distribution pattern reflects their putative functional connection. All the most concerning variants carry one or (typically) more mutations in each of these three domains. The NTD domain accumulates the highest number of convergent mutations. Strikingly, all these AA residues, despite being located apart from each other in the S-protein primary structure, form perfectly co-localized patches, literally delineating the site of evolutionary pressure for optimization. We speculate that this adaption goes beyond a simple immune evasion or antigenic minimalization (Venkatakrishnan et al., 2021), which is unlikely to explain this level of convergence in NTD evolution. Instead, we suppose that NTD mutations might modulate some functional properties of this domain, which unfortunately remains experimentally understudied. The described NTD functions include controlling S-protein conformation (Li et al., 2021), interaction with host surface sialosides (Awasthi et al., 2020) and, most recently, binding to alternative entry receptors (Zhu et al., 2021). The second hot spot – the RBD domain shows a strong accumulation of mutations at the binding interface with the ACE2 receptor (see below for a dedicated paragraph). The function of the Furin cleavage site is well established (Xia et al., 2020; Johnson et al., 2021) as well as the relevance of the convergent Q677H mutation (Hodcroft et al., 2021). The impact of additional mutations in this region remains to be analyzed, but their positive fitness effect due to faster S-protein processing can be expected. As with the NTD, the Furin cleavage site AAs under convergent evolution are in direct contact with each other. The only convergent mutations not within these three domains are above residue 700, and their parallelization score is low (**Figure 2A**, inset d). It should be noted that these convergent mutations evolved on the background of a mostly naïve host. Conversely, the search for previously unseen convergent evolution in delta and omicron variants diversity since mid-2021 gained only three new significant mutations. A222V appeared at least three times among delta lineages, T95I emerged multiple times in delta and is present in omicron, and finally R346K was detected in omicron lineages.

Combinations of Convergent Mutations Suggest Epistatic Relations

The spatial clustering of AAs under convergent evolution creates room for synergic effects between mutations, as was observed for the couple E484K/N501Y and proposed for Q498R/N501Y
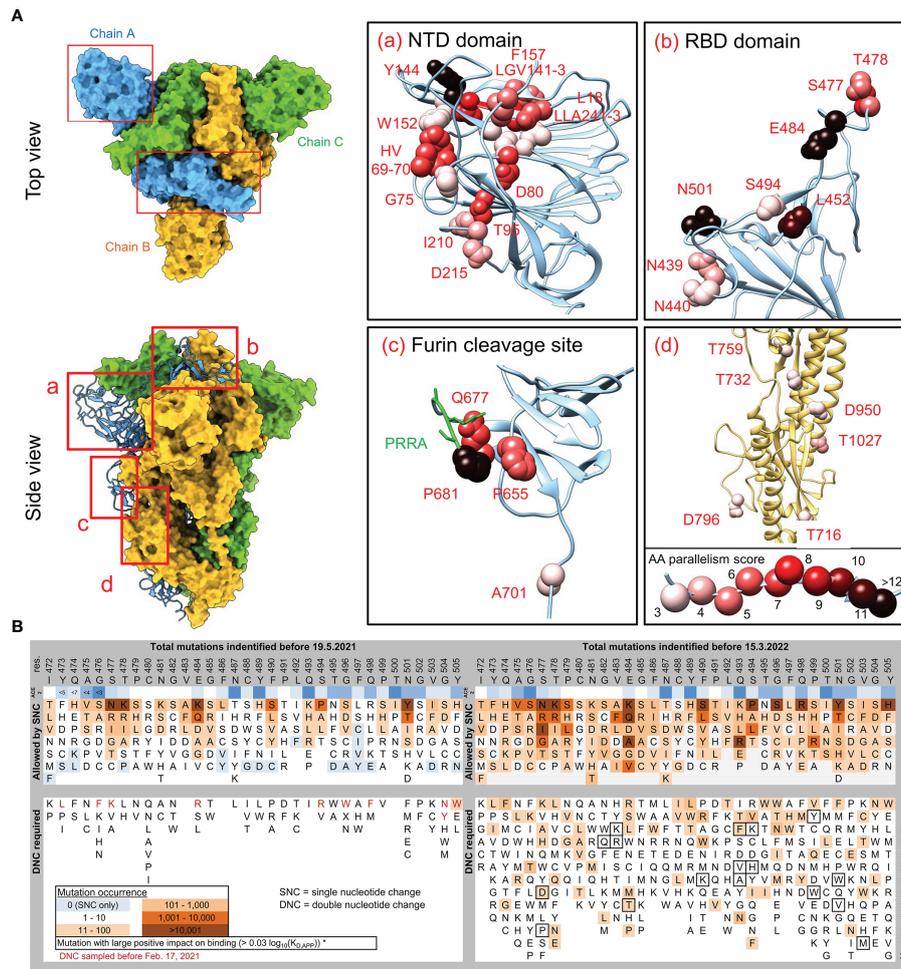
Original SARS-CoV-2 S-protein residue

PANGO lineage	Alternative lineage identifier	Geographic prominence / origin	L18	HV69-70	G75	D80	T95	D138	LGV141-3	Y144	W152	F157	I210	D215	LLA241-3	N439	N440	L452	S477	T478	E484	S494	N501	H655	Q677	P681	A701	T716	T732	D796	T859	D950	T1027	Lineage parallelism score
AV.1	VUI-21MAY-01	U.K.			G		I	D*	Δ							K			K							H								68
B.1.620		Europe		Δ				Δ							Δ				K							H							I	64
B.1.617.1	kappa	India						I#	D*#	S								R#			Q					R#								55
B.1.621		Colombia					I			S									K			Y				H					N			55
	A.VOI.V2	Central Africa				Y						N	G							R		Y			Y	H								55
P.3	theta	Philippines						Δ							Δ				K			Y				H								54
B.1.1.318	VUI-21FEB-04	Central Africa					I												K							H				H				54
B.1.617.2	delta	India						I#	D*#		Δ							R#		K					R#						N			52
B.1.1.7	alpha	Europe		Δ					Δ														Y			H		I						52
B.1.351	beta	South Africa	F			A							G	Δ					K			Y					V							51
P.1/4	gamma	Brazil	F					Y											K			Y	Y								Y		I	50
A.27	HMN.19B	France	F													R						Y	Y	H					Y				45
C.36.3		Egypt		Δ					R						R			K					H											44
B.1.525	eta	Central Africa		Δ				Δ							K								H										44
B.1.526.1		U.S.			G			Δ		S					R									Y				N	H				43
B.1.604		Angola						Δ							R						Y			R					Y				42
AT.1		Russia					Δ	Δ			G						K																41
N.10		Brazil					Δ	Δ	V								K																41
A.29		U.K.		Δ				Δ												Y	Y												33
B.1.616		France					V*	Δ			G										Y												31
B.1.623		U.S., Lesser Antilles														P	Y						H										31
B.1.1.523		Russia							Δ							K	P															I	28
A.28		Europe		Δ																	T	Y				V							26
B.1.526	iota	U.S., Ecuador					I#											K						V									25
B.1.1.519		Mexico															K							H		A							23
B.1.160.24		Sweden						Δ									N																22
B.1.575		Dominican																	P				H		I								21
A.23.1		East Africa							L														R										20
R.1		Japan							L						K																		19
B.1.177.80		Scandinavia	F#					Δ																									19
A.2.5		Central America						Δ								R																	18
AU.2/AU.3		Oceania													K								R										18
C.37		South America			V											Q													N				17
N.9		Brazil														K																	16
B.1.243		North America	F																				H										16
B.1.1.420		Senegal										K								H													15
B.1.214.3		Europe					I							K										I									14
B.1.429/427	epsilon	U.S.							C			R																					14
L.3		Central Africa			R							R																					14
B.1.1.317		Russia					Y						N																				14
B.1.1.413		Europe					Y						N																				14
B.1.526.2		U.S.					I#						N																				13
AP.1		U.K.													Y																		13
B.1.517		U.S.													T																		13
C.2.1		Aruba													T																		13
A.21		West Africa									R																						11
B.1.619		Cameroon							T			K																				I	11
C.16		South Africa									R																						11
B.1.1.316		U.S.																				H			S							10
B.1.258		Europe		Δ							K																						10
B.1.127		Denmark											N																				8
B.1.1.375		Mozambique											N																				8
B.1.160		global											N																				8
B.1.36.9		U.K.							Δ																N							7
B.1.596		U.S.																				P										7
B.1.1.70		Europe																				H										7
B.1.375		U.S.		Δ																												7
B.1.177.30		U.K.				Y																										7
B.1.1.417		Canada				Y																										7
B.1.367		Europe				Y																										7
B.1.1.397		Russia						Y																									6
B.1.1.333		Norway						Y																									6
B.1.1.224		Denmark						Y																									6
B.1.177		Europe	F#																														5
B.1.36.1		Turkey																K															5
B.1.1.219		Europe							L																								5
B.1.36.29		India										K																					3
B.1.1.222		Mexico																								A							3
B.1.524		SE Asia																									V						3
B.1.36.18		Canada			R																												3
AA parallelism score			5	7	3	7	6	6	7	14	3	5	4	4	3	3	3	11	8	5	16	3	13	6	7	15	3	3	3	3	3	3	4	

Functional domain			Within-lineage mutation frequency			
NTD	RBD	Furin cleavage site	100-75%	75-50%	50-25%	

FIGURE 1 | Convergent evolution of SARS-CoV-2 S-protein AA residues among different lineages (autumn 2020 – spring 2021). All S-protein AA mutations present in SARS-CoV-2 VOC lineages (alpha, beta, gamma, delta) were collected. The phylogenetic pattern of mutations at each of these AA positions was assessed visually in the representative global subsample of SARS-CoV-2 genomes (https://nextstrain.org/ncov/global) (Hadfield et al., 2018). NextStrain was scanned repeatedly between March to June 2021; lineages were confirmed to show a recent localized rise to high frequency (indicative of selective advantage). All additional S-protein AA mutations present in these lineages were subjected to the same procedure as above. After several iterations, a final set of S-protein positions that experienced ≥3 independent mutations was established. Information about the distribution and within-lineage frequency of S-protein mutations, and the spatiotemporal characterization of SARS-CoV-2 lineages was retrieved from (Tsueng et al., 2022) (15 June 2021). The AA parallelism score was established for all convergently mutated AAs, by summing their independent emergence events. The lineage parallelism score was calculated by summing the parallelism scores of S-protein AAs mutated in individual lineages. *AA substitutions for G142; Mutations inherited from the same progenitor are considered a single evolutionary origin in parallelism score calculations.

FIGURE 2 | Localization of convergent mutations in SARS-CoV-2 S-protein structure, breakdown and progression of RBD AA mutations. **(A)** The S-protein parts missing in the crystal structure (PDB ID: 6zge) were modelled by the Modeller suite implemented in UCSF Chimera 1.13.1 (Webb and Sali, 2016). Residues under convergent evolution are depicted in spheres representation and colored according to their parallelism scores (**Figure 1**). Inset d) shows the region covering a portion of heptad repeat, central helix, and β-hairpin domains. Green residues in inset c) highlight the Furin cleavage site. **(B)** The mutational sequence space of RBD binding interface residues 472 – 505, and its coverage by mutations present in the GISAID database. All possible SNC (single-nucleotide change) AA substitutions are depicted. For TNC (two-nucleotide change) substitutions, only the subset of AA substitutions that were sampled in GISAID are depicted (in background color scale according to the legend), together with substitutions with a positive binding impact (frame). The AA position is invariant to its later occurrence to highlight differences. Deep-mutational scanning $\Delta Log_{10}(K_{D, App})$ values were extracted from https://jbloomlab.github.io/SARS-CoV-2-RBD_DMS/ (Starr et al., 2020). ACE2 lane depicts residue distances in Å from the ACE2 receptor.

(Zahradník et al., 2021) that later emerged in omicron lineages. In addition, potential inter-domain relations in mutational appearance can be tentatively inferred. For example, as a whole, NTD convergent mutations predominate in lineages with mutated RBD. Furthermore, some NTD convergent mutations (e.g. T95I, ΔY144, D215G) accumulate in lineages with high parallelism scores and are never present alone (**Figure 1**). This indicates that the S-protein adaptive mutations emerge in a sequential manner. The precise order of mutation appearance is difficult to analyze since the lineages of concern with a high mutational load typically appear "out of nowhere" without a documented intermediate state. For the SARS-CoV-2 research, this means

that, ideally, S-protein mutations should be experimentally evaluated in (both intra- and inter-domain) combinations.

Exploring Double Nucleotide Changes by SARS-CoV-2 RBD

The RBD domain is the focus of our understanding of SARS-CoV-2 cell entry and neutralizing antibody evasion (Starr et al., 2020; Dejnirattisai et al., 2022), and its evolution thus deserves special attention. We generated an *in silico* library of all RBD AA mutations which can be achieved with a single nucleotide change (SNC) and two-nucleotide changes (TNC) of the original codons, and analyzed their global occurrence. Rapid

progression can be seen as all of the possible SNC-dependent changes had been detected in early 2021 among the >1.2 million genomic sequences in GISAID. As the likelihood of two successive nucleotide changes of a single codon is much lower, AA mutations resulting from TNC appear among sampled genomes at much lower frequencies, yet their repertoire and numbers are growing over time and soon are likely to reach complete exhaustion (**Figure 2B**). The first TNC detected among VOC/VOI variants is omicron´s S371L mutation, with others predicted to follow. Some of the changes conferred by TNC are predicted to display a much tighter binding to ACE2 (**Figure 2B**) and thus represent a potential epidemiological risk. We can assume that the occurrence of epistatic mutations is also expanding, as seen for Q498R in the omicron. Taken together, the potential sequence space to be sampled in the future by the RBD to achieve an evolutionary advantage remains large and we will see a rise of previously unseen mutations.

Both Receptor Binding and Immune Evasion Are Shaping the RBD Evolution

Most of the AA changes predicted to have a positive effect on ACE2 receptor binding [$Log10(K_{D,app}) > 0.03$] (Starr et al., 2020) are present at significant frequencies among GISAD genomes (**Figure 2B**), reinforcing improved receptor binding as an importing contributor to the adaptive fitness gain. Mutations into positively charged residues are favored by electro-complementarity between RBD and ACE2 (Zahradník et al., 2021). Among the convergent RBD mutations, this concerns N439K, N440K, L452R, T478R/K and E484K (**Figure 1**). Expectedly, the selection of most mutations that negatively affect receptor binding is disfavored in RBD evolution (**Figure 2B**). Our predictions of SARS-CoV-2 evolution often involve immunocompromised patients, treated or non-treated with convalescent plasma or neutralizing antibodies (Tarhini et al., 2021; Truong et al., 2021), yet it is not clear how these two phenomena (global vs. within-patient virus evolution/diversification) mirror each other. While several RBD mutations have been selected during prolonged infection within the same immunocompromised patient (Choi et al., 2020; Clark et al., 2021), others were not e.g. the combination of E484K/N501Y, which increases the binding affinity to ACE2 receptor (Zahradník et al., 2021). In contrast, the NTD deletion Δ141-143 has evolved in multiple patients (Clark et al., 2021; Kemp et al., 2021) and, like ΔY144 (Chen et al., 2021), was clearly selected in response to the host environment. Some RBD mutations may increase viral fitness by immune evasion mechanisms other than reduced antibody neutralization, e.g. reduced MHC presentation [L452R (Motozono et al., 2021)]. It is thus evident that every mutation contributes to the viral fitness by different parameters and mechanisms, which need to be evaluated systematically.

Global Selective Sweep Since Early 2021

The global evolutionary dynamics of SARS-CoV-2 has changed fundamentally since early 2021. The early common pattern (competition of multiple independent lineages on the background of the naïve population) gave way to the global selective sweep of delta VOC. Concomitantly, the rapid vaccination rate in combination with post-infection immunity made the virus face a significantly more immunized population.

The second half of 2021 was characterized by the delta diversification into numerous competing sublineages. Finally, in November, omicron emerged from an unknown, presumably "near" wildtype SARS-CoV-2 (B.1.1) and has been globally spreading at an unprecedented rate. Omicron carries total 13 mutations in S-protein which were convergent in the pre-omicron era (HV69-70Δ, T95I, G142D, Y144Δ, N440K, S477N, T478K, E484A, N501Y, H655Y, P681H, D796Y, D950N), double of the maximum convergent mutations observed in any other VOCs so far. In this sense, omicron is an excellent example how these residues were built up again to form omicron, as no other variant had over seven of them (**Figure 1**) . In addition, the breakdown of RBD mutations prior of the omicron era showed omicron residues to be among the most prominent ones. The only exception is the mutation Q498R due to the already mentioned cooperativity to N501Y. Our work thus shows that SARS-CoV-2 early evolution was towards human adaptation which, despite remaining closed and unrepeatable, presaged future developments.

The extent and progression of convergence in S-protein evolution tentatively hints at the possibility that it might approach a final state of optimization to the novel, human host (Burioni and Topol, 2021). However, with the large mutation space available by rarer TNC and epistatic mutations, this is far from certain. Still, with the predictability and repeatability of S-protein evolution taken into consideration in vaccine design (even in omicron), the protection for the global population from the landscape of viral variants of current and near-future significance can be maximized. Clearly, mass vaccination and/or previous infections are altering the selective pressure and thus reshaping viral evolution. This phenomenon is already playing a role at position 484, where lysine is the optimal residue, but omicron variants is characteristic with E484A. It should be noted that omicron emerged in South Africa regions previously affected by E484K-bearing variants (gamma and beta). Also in omicron, the highly increased binding affinity of the Q498R/N501Y double-mutant was used to insert a slew of other mutations, that while reducing binding affinity promoted immune evasion (Dejnirattisai et al., 2022). In summary, taking S-protein convergent evolution into consideration could have provided us with the much-needed time to design and test broad-range VOC-effective vaccines in advance of the real developments. The need for global monitoring of SARS-CoV-2 evolution will, for some time, certainly remain a top priority.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: GISAID.

AUTHOR CONTRIBUTIONS

JZ and GS conceived the project; JZ, JN, and GS performed experiments; JZ, JN, and GS wrote the manuscript. All authors contributed to the article and approved the submitted version.

REFERENCES

Awasthi, M., Gulati, S., Sarkar, D. P., Tiwari, S., Kateriya, S., Ranjan, P., et al. (2020). The Sialoside-Binding Pocket of SARS-CoV-2 Spike Glycoprotein Structurally Resembles MERS-CoV. *Viruses* 12 (9), 909. doi: 10.3390/v12090909

Burioni, R., and Topol, E. J. (2021). Has SARS-CoV-2 Reached Peak Fitness? *Nat. Med.* 27, 1323–1324. doi: 10.1038/s41591-021-01421-7

Chen, L., Zody, M. C., Germanio, C. D., Martinelli, R., Mediavilla, J. R., Cunningham, M. H., et al. (2021). Emergence of Multiple SARS-CoV-2 Antibody Escape Variants in an Immunocompromised Host Undergoing Convalescent Plasma Treatment. *mSphere* 6 (4), e00480–e00421. doi: 10.1128/mSphere.00480-21

Cherian, S., Potdar, V., Jadhav, S., Yadav, P., Gupta, N., Das, M., et al. (2021). SARS-CoV-2 Spike Mutations, L452R, T478K, E484Q and P681R, in the Second Wave of COVID-19 in Maharashtra, India. *Microorganisms* 9 (7), 1542. doi: 10.3390/microorganisms9071542

Choi, B., Choudhary, M. C., Regan, J., Sparks, J. A., Padera, R. F., Qiu, X., et al. (2020). Persistence and Evolution of SARS-CoV-2 in an Immunocompromised Host. *N. Engl. J. Med.* 383 (23), 2291–2293. doi: 10.1056/NEJMc2031364

Clark, S. A., Clark, L. E., Pan, J., Coscia, A., McKay, L. G.A., Shankar, S., et al. (2021). SARS-CoV-2 Evolution in an Immunocompromised Host Reveals Shared Neutralization Escape Mechanisms. *Cell* 184 (10), 2605–2617.e2618. doi: 10.1016/j.cell.2021.03.027

Dejnirattisai, W., Huo, J., Zhou, D., Zahradník, J., Supasa, P., Liu, C., et al. (2022). SARS-CoV-2 Omicron-B.1.1.529 Leads to Widespread Escape From Neutralizing Antibody Responses. *Cell* 185 (3), 467–484.e415. doi: 10.1016/j.cell.2021.12.046

De Maio, N., Walker, C. R., Turakhia, Y., Lanfear, R., Corbett-Detig, R., Goldman, N., et al. (2021). Mutation Rates and Selection on Synonymous Mutations in SARS-CoV-2. *Genome Biol. Evol.* 13 (5), evab087. doi: 10.1093/gbe/evab087

Hadfield, J., Megill, C., Bell, S. M., Huddleston, J., Potter, B., Callender, C., et al. (2018). Nextstrain: Real-Time Tracking of Pathogen Evolution. *Bioinformatics* 34 (23), 4121–4123. doi: 10.1093/bioinformatics/bty407

Hodcroft, E. B., Domman, D. B., Snyder, D. J., Oguntuyo, K., Van Diest, M., Densmore, K. H., et al. (2021). Emergence in Late 2020 of Multiple Lineages of SARS-CoV-2 Spike Protein Variants Affecting Amino Acid Position 677. *medRxiv*, 2021.2002.2012.21251658. doi: 10.1101/2021.02.12.21251658

Johnson, B. A., Xie, X., Bailey, A. L., Kalveram, B., Lokugamage, K. G., Muruato, A., et al. (2021). Loss of Furin Cleavage Site Attenuates SARS-CoV-2 Pathogenesis. *Nature* 591 (7849), 293–299. doi: 10.1038/s41586-021-03237-4

Kemp, S. A., Collier, D. A., Datir, R. P., Ferreira, I. A.T.M., Gayed, S., Jahun, A., et al. (2021). SARS-CoV-2 Evolution During Treatment of Chronic Infection. *Nature* 592 (7853), 277–282. doi: 10.1038/s41586-021-03291-y

Lemmermann, N. A., Lieb, B., Laufs, T., Renzaho, A., Runkel, S., Kohnen, W., et al. (2021). SARS-CoV-2 Genome Surveillance in Mainz, Germany, Reveals Convergent Origin of the N501Y Spike Mutation in a Hospital Setting. *medRxiv*, 2021.2002.2011.21251324. doi: 10.1101/2021.02.11.21251324

Li, Y., Wang, T., Zhang, J., Shao, B., Gong, H., Wang, Y., et al. (2021). Exploring the Regulatory Function of the N-Terminal Domain of SARS-CoV-2 Spike Protein Through Molecular Dynamics Simulation. *Adv. Theory Simul.* 4 (10), 2100152. doi: 10.1002/adts.202100152

Majumdar, P., and Niyogi, S. (2021). SARS-CoV-2 Mutations: The Biological Trackway Towards Viral Fitness. *Epidemiol. Infect.* 149, e110. doi: 10.1017/S0950268821001060

Martin, D. P., Weaver, S., Tegally, H., San, J. E., Shank, S. D., Wilkinson, E., et al. (2021). The Emergence and Ongoing Convergent Evolution of the SARS-CoV-2 N501Y Lineages. *Cell* 184 (20), 5189–5200.e5187. doi: 10.1016/j.cell.2021.09.003

Motozono, C., Toyoda, M., Zahradnik, J., Saito, A., Nasser, H., Tan, T. S., et al. (2021). SARS-CoV-2 Spike L452R Variant Evades Cellular Immunity and Increases Infectivity. *Cell Host Microbe*. doi: 10.1016/j.chom.2021.06.006

Ramazzotti, D., Angaroni, F., Maspero, D., Gambacorti-Passerini, C., Antoniotti, M., Graudenzi, A., et al. (2021). VERSO: A Comprehensive Framework for the Inference of Robust Phylogenies and the Quantification of Intra-Host Genomic Diversity of Viral Samples. *Patterns* 2 (3), 100212. doi: 10.1016/j.patter.2021.100212

Shu, Y., and McCauley, J. (2017). GISAID: Global Initiative on Sharing All Influenza Data - From Vision to Reality. *Euro. Surveill.* 22 (13), 30494. doi: 10.2807/1560-7917.ES.2017.22.13.30494

Starr, T. N., Greaney, A. J., Hilton, S. K., Ellis, D., Crawford, K. H.D., Dingens, A. S., et al. (2020). Deep Mutational Scanning of SARS-CoV-2 Receptor Binding Domain Reveals Constraints on Folding and ACE2 Binding. *Cell* 182 (5), 1295–1310.e1220. doi: 10.1016/j.cell.2020.08.012

Tarhini, H., Recoing, A., Bridier-Nahmias, A., Rahi, M., Lambert, C., Martres, P., et al. (2021). Long Term SARS-CoV-2 Infectiousness Among Three Immunocompromised Patients: From Prolonged Viral Shedding to SARS-CoV-2 Superinfection. *J. Infect. Dis* 223(9):1522–1527. doi: 10.1093/infdis/jiab075

Truong, T. T., Ryutov, A., Pandey, U., Yee, R., Goldberg, L., Bhojwani, D., et al. (2021). Increased Viral Variants in Children and Young Adults With Impaired Humoral Immunity and Persistent SARS-CoV-2 Infection: A Consecutive Case Series. *EBioMedicine* 67, 103355. doi: 10.1016/j.ebiom.2021.103355

Tsueng, G., Mullen, J., Alkuzweny, M., Cano, M., Rush, B., Haag, E., et al. (2022). Outbreak.info Research Library: A Standardized, Searchable Platform to Discover and Explore COVID-19 Resources and Data. *bioRxiv*, 2022.2001.2020.477133. doi: 2022.2001.2020.477133

Venkatakrishnan, A. J., Anand, P., Lenehan, P., Ghosh, P., Suratekar, R., Siroha, A., et al. (2021). Antigenic Minimalism of SARS-CoV-2 is Linked to Surges in COVID-19 Community Transmission and Vaccine Breakthrough Infections. *medRxiv*, 2021.2005.2023.21257668. doi: 10.1101/2021.05.23.21257668

Volz, E., Hill, V., McCrone, J. T., Price, A., Jorgensen, D., O'Toole, Á., et al. (2021). Evaluating the Effects of SARS-CoV-2 Spike Mutation D614G on Transmissibility and Pathogenicity. *Cell* 184 (1), 64–75.e11. doi: 10.1016/j.cell.2020.11.020

Webb, B., and Sali, A. (2016). Comparative Protein Structure Modeling Using MODELLER. *Curr. Protoc. Bioinf.* 54, 5.6.1–5.6.37. doi: 10.1002/cpbi.3

Xia, S., Lan, Q., Su, S., Wang, X., Xu, W., Liu, Z., et al. (2020). The Role of Furin Cleavage Site in SARS-CoV-2 Spike Protein-Mediated Membrane Fusion in the Presence or Absence of Trypsin. *Signal Transduct. Targeted Ther.* 5 (1), 92. doi: 10.1038/s41392-020-0184-0

Zahradník, J., Marciano, S., Shemesh, M., Zoler, E., Harari, D., Chiaravalli, J., et al. (2021). SARS-CoV-2 Variant Prediction and Antiviral Drug Design are Enabled by RBD *In Vitro* Evolution. *Nat. Microbiol.* 6 (9), 1188–1198. doi: 10.1038/s41564-021-00954-4

Zhou, H.-Y., Ji, C.-Y., Fan, H., Han, N., Li, X.-F., Wu, A., et al. (2021). Convergent Evolution of SARS-CoV-2 in Human and Animals. *Protein Cell.* 12 (11), 832–835. doi: 10.1007/s13238-021-00847-6

Zhu, S., Liu, Y., Zhou, Z., Zhang, Z., Xiao, X., Liu, Z., et al. (2021). Genome-Wide CRISPR Activation Screen Identifies Candidate Receptors for SARS-CoV-2 Entry. *Sci. China Life Sci.* 1–17. doi: 10.1007/s11427-021-1990-5