Comparison of Sars-Cov-2 Mutation Patterns in Three Closed Populations Over Time

Mario Vilela[1], Deena Mikbel[1*], Amir Hamidi[1], Lora Benoit[2], Ali Mohseni[3 †]*

[1]University of California, Davis, [2]California Health Sciences University College of Osteopathic Medicine, [3]Persian American Civil Action Network

* Contributed equally to this project, † Presenting Author

INTRODUCTION

The ongoing SARS-CoV-2 pandemic has caused significant disease burden worldwide. Since the emergence of the reference strain in 2019, the virus continues to undergo mutations that influence infectivity/ transmission and evasion of the innate and adaptive immune response. To better understand the role of the immune system in evolving the mutations, we have characterized mutations at ~ 6 and ~ 12 months into the Pandemic using randomly selected mutant strains derived from 3 closed populations (Iceland, Iran, and Japan) and compared them against the reference strain and three widely circulating variants of concern (VOCs). For each mutant strain analyzed, we have characterized the mutations occurring in each ORF, assessed the predictability of the mutations, and assessed potential changes in immunogenicity in the most antigenic targets, specifically the S, M, and N proteins using bioinformatic platforms. At both timepoints, we found that mutations occurring in sequences from Iran differed significantly from those occurring in Iceland and Japan, specifically demonstrating increased variability in mutation occurrence, number of deletions, NS/S mutation ratios, and in the mutations affecting B cell/antibody binding and T cell/HLA binding epitopes. Although the % blastN identify for S protein was similar for all 3 countries at both timepoints, the Iranian sequences had significantly more mutations affecting antibody binding, with potential implications for current vaccine efficacy and monoclonal antibody interventions. All three demonstrated significant and distinct shifts in Pango Lineages over time, underscoring the dynamic nature of the mutations in each population assessed. These findings are discussed within the context of regional immunogenetics and help explain the selection pressures that are driving the genetic diversification of SARS-CoV-2 with application to prophylaxis.

METHODS

To assess the mutations, ~22 whole genome-length mutant strain sequences (>29,400 informative nucleotides in length) collected during the summer and winter of 2020 from three closed populations (Iceland, Japan, and Iran) were randomly selected from the Genbank and GISAID databases. Each mutant sequence was compared against the reference strain (NC_045512.2) sequence at both the nucleotide and amino acid level. Comparisons were performed using the alignment tool **HTTPS://WWW.EBI.AC.UK/TOOLS/PSA/EMBOSS_NEEDLE**. The results were then catalogued for each ORF, and individual molecular assessments were performed. To assess whether mutations might be influencing the adaptive immune response, immunodominant epitopes were assessed in each mutant variant. As the most wildly immunogenic epitopes have been reported to occur in structural proteins S, M, and N (3,4,5), we thus restricted our analysis of immunodominant epitopes to these 3 targets. For B cells, mutant amino acid sequences for S, M, and N were compared against previously reported immunodominant antibody epitopes (3,4,5) as well as sites identified in GISAID. For T cells, we ascertained the most frequent HLA-I and HLA-II alleles for each closed population using data integrated from **WWW.ALLELEFREQUENCIES.NET**. To define the most immunodominant epitope for each HLA allele, then ran the reference strain and corresponding individual mutation sequences for S, M and N through **WWW.CBS.DTU.DK/SERVICES/NETMHCPAN** and **WWW.CBS.DTU.DK/SERVICES/NETMHCIIPAN** to determine the highest binding epitope for each HLAs defined for each population. For mutant strains possessing an amino acid mutation in each of these immunodominant epitopes, we determined the binding affinity with and without the mutation to determine the predicted impact on immune recognition. In addition to comparing these regionally defined sequences, we analyzed the sequences of 3 wildly circulating Variants of Concern (VOCs), specifically the UK variant (B.1.1.7; EPI_ISL_601443), the Brazilian Variant (P.1 or formally B.1.1.28.1; EPI_ISL_833174) and the South African Variant (B.1.351; EPI_ISL_736980). Comparisons were made directly against the reference strain and each VOC was also assessed against each of the 3 closed populations in terms of HLA-binding.

RESULTS AND CONCLUSIONS

Mutational Analysis (Chart 1): Mutation rate is a critical parameter for understanding viral evolution. **As for nucleotide mutations in SARS-CoV-2, a strong mutational biases marked by 1) non-synonymous to synonymous nucleotide mutation (NS/S) ratio of 1.88 and a high proportion of A→G changes and C→T (U) changes (roughly 50%) due to the activity of host RNA-editing enzymes has previously been reported and are therefore predicted.** This latter property denotes a viral evasion strategy via the "silencing" of antigenic nucleic acids so that they are poorly recognized by viral sensors. The sequences from Iran consistently demonstrated much more variability in the number of NT and AA substitutions compared to the sequences from the other 2 regions (20.4+/-37.4 and 21.9+/-10.0 for NTs, and 13.3+/-20.2 and 13.1+/- 4.7 for AA at the 2 time points), likewise, they demonstrated a higher proportion of deletions as well, most notably at the second time point (3/23 sequences at the first time point, and 10/12 sequence at the second time point). Moreover, the NS/S ratio was higher in the mutant sequences from Iran (3.9 and 3.5), above the predicted ratio. This latter finding was also mirrored in the South African Variant of Concern (VOC), which had an unusually high NS/S ratio of 9.5. The frequency of C-> T (U) mutations were lower in the mutant sequences from Iran (although not significantly), compared to those from Japan and Iceland, as was the case in the 3 VOCs from UK, South Africa and Brazil. With the exception of VOC South Africa, cases, the open reading frames (ORFs) with the most mutations followed the expected pattern of ORf1ab, S and N, though not necessarily in that order. Looking more carefully at the lineage changes over time, Iran sequences shifted from a B.4 (Iranian Export) dominant line -> B.1.36, Icelandic sequences shifted from B.1.1.232 -> B.1.1.177, and Japanese sequences shifted from B.1.1 -> B.1.1.214. At the second time point, Iran reported 1/12 UK VOC and Japan reported 1/22 UK VOCs, while Iceland still did not have any VOC in the random cohort assessed.

Chart 1: Mutation Summary:

Data set	Strains	ave. # Nt mutations	Nt insertion events	Nt deletion events	ave. # AA substitutions	average NS/S ratio	Top 3 mutated ORFs
VOC	VOC UK	45	0	1	17	1.7	Orf1ab> S> N
	VOC SA	39	0	1	19	9.5	S> Orf1ab> Orf3a
	VOC BR	40	0	1	21	2.1	Orf1ab> S> N
6 months	Iceland (22)	16.2 +/-3.1	0	1 (out of 22)	10.4 +/-2.4	2.2 +/-0.9	Orf1ab> N > S
	Iran (23)	20.4+/-37.4	0	3 (out of 23)	13.3 +/- 20.2	3.9+/- 4.3	Orf1ab> S> N
	Japan (22)	11.5 +/- 2.4	1 (out of 22)	1 (out of 22)	8.1 +/- 2.9	1.9 +/- 0.7	Orf1ab > N > S
12 months	Iceland (21)	16.9 +/- 2.8	0	1 (out of 21)	9.1 +/-1.8	1.4 +/- 0.6	Orf1ab> S> N
	Iran (12)	21.9 +/- 10.0	0	10 (12)	13.1 +/- 4.7	3.5 +/- 3.4	Orf1ab> S> N
	Japan (22)	17.6 +/- 6.8	0	1 (out of 22)	10.5 +/-3.3	1.8 +/- 0.8	Orf1ab> N > S

Data set	Strains	ave. C-> T (U) mutations # (%)	ave. A-> G mutations # (%)	blastN % S identity	% D614G+	Dominant Pango Lineage	VOC Occurrence
VOC	VOC UK	10 (22.2%)	2 (4.4%)	99.58%	yes	B.1.1.7	
	VOC SA	9 (23.1%)	4 (10.3%)	99.56%	yes	B.1.351	
	VOC BR	10 (25%)	3 (7.5%)	99.69%	yes	P.1 (B.1.1.28.1)	
6 months	Iceland (22)	4.6 (28.8%)	1.5 (10.3%)	99.95%	100%	B.1.1.232 (63.6%)	none
	Iran (23)	3.8 (29.5%)	2.0 (9.4%)	99.92%	43.50%	B.4 (56.5%)	none
	Japan (22)	5.4 (46.6%)	1.3 (11.2%)	99.96%	100%	B.1.1 (54.5%)	none
12 months	Iceland (21)	7.2 (43.5%)	1.6 (9.2%)	99.93%	100%	B.1.177 (38.1%)	none
	Iran (12)	8.1 (39.3%)	2.6 (14.1%)	99.88%	100%	B.1.36 (41.7%)	1 UK VOC
	Japan (22)	8.0 (47.7%)	1.6 (9.3%)	99.94%	100%	B.1.1.214 (54.5%)	1 UK VOC

B cell Epitopes: We assessed mutations occurring in B cell epitopes derived from S protein, 2 epitopes from N protein (all major immunogens for SARS-CoV-2)) (Chart 2). We analyzed previously reported SARS-CoV-2 immunodominant B cell epitopes (3,4,5), GISAID-reported antibody-binding sites, and as well as VOC spike protein mutations (S 144, 417, 484, 501). Change in binding affinity was calculated using the bioinformatics tool **HTTP://TOOLS.IEDB.ORG/BCELL** by comparing the affinity of the reference sequence against that of the mutant sequence. At the first time point (6 months), only mutants from Iran 39.1% of sequences assessed demonstrated mutations in antibody-binding site –most of which reduced antibody binding, those from Iceland and Japan were unaffected. At the second time point (12 months), mutants from all 3 regions now possessed mutations in antibody-binding sites, though the frequency in the Iranian sequences was much higher compared to the mutants from the other 2 regions (33.3% of sequences vs < 5% of sequences). This finding indicates that the mutant strains circulating in Iran appear to be selecting against the antibody response and moreover, are potentially more resistant to the neutralizing effects of vaccine-induced antibodies/immunity.

Chart 2: B Cell Immunodominant Epitopes (S,M,N) Event Summary:

Data set		AA Δs in S epitopes	AA Δs in M epitopes	AA Δs in N epitopes	AA Δs in Ab sites per GISAID*
VOC	VOC UK	1 -mild ↓	0	0	S 144,501
	VOC SA	0	0	0	S 417, 501, 484
	VOC BR	1 -mild ↓	0	N	S 417, 501, 484
Iceland	6 months (22)	0	0	0	0%
	12 months (21)	1 -mild ↓	0	2 -modest ↓	4.8% (S)
Iran	6 months (23)	4 mild ↓	0	0	39.1% (S and N)
	12 months (12)	0	0	1 -modest ↓, 1 =	33.3% (S only)
Japan	6 months (22)	0	0	0	0%
	12 months (22)	1 -mild ↓	0	0	4.5% (S only)

* Indicated by GISAID, other epitope mutations are published

HLAs and HLA-binding Peptides /T cell Epitopes: The most frequent HLAs were derived from Allele Frequency Net Database by integrating data sets reporting high resolution typing. For Iceland, the HLAs were extrapolated from data sets reported for Ireland, Norway, and Sweden. Epitopes from the reference strain (with the highest HLA-binding affinity for each HLA allele) and each corresponding mutant strain were compared at 6 months and 12 months (**Chart 3**).

At the early time point (~6 months): sequences from Iran had the most mutations in T cell immunodominant epitopes. Specifically, Iran had 6/23 mutations, 3 occurring in HLA-I-binding epitopes of S and N and 3 occurring in HLA-II binding epitopes of S and N –most of which effected mild reductions in HLA binding. Iceland had 1/22 mutations in an HLA-binding epitope of N. Both Iran and Iceland possessed 1 mutation in the N protein which severely reduced binding (~90%) to HLA-I. There were only 2/22 mutations in HLA-II-binding epitopes in sequences from Japan, both of which mildly effecting affinity.

At the later time point (~12 months): Iran had 3/12 mutations in immunodominant epitope of M protein that results in a mild reduction of affinity to HLA-II. This mutation is significant in that M protein is a relatively conserved protein. Iceland had 9/21 mutations in immunodominant epitope of N protein that results in a mild reduction of affinity to HLA-II. Japan had 6/21 mutations in immunodominant epitope of N protein and 1 /21 mutations in immunodominant epitope of S protein, all of which resulted in a modest reduction of affinity. Importantly, no HLA-I immunodominant epitopes were noted in the sequences analyzed.

Overall, early on in Iran, there were many mutations that affected HLA binding (Class I and II) and the frequency did not change over time. **In comparison, there were very few mutations in sequences from Iceland and Japan early on, though this changed significantly at the later time point, implying that over time, the virus was undergoing mutations in immunodominant epitopes that influenced HLA-II binding, indicating that the virus was evolving over time in these 2 populations to escape CD4+ T helper cells. This is significant in that CD4+ T cells critically control B cell and CD8+ Cytotoxic T cell responses to the virus.**

Chart 3: T cell Immunodominant Epitopes (S,M, N) Events Summary:

Location	Strains		HLA- Class 1 (HLA-A, HLA-B, HLA-C)			HLA- Class 2 (HLA-DR, HLA-DQ, HLA-DP)		
			S epitopes	M epitopes	N epitopes	S epitopes	M epitopes	N epitopes
Iceland	6 months (22 samples)	VOC UK	0	0	0	0	0	0
		VOC SA	0	0	0	0	0	0
		VOC BR	0	0	0	0	0	0
		Iceland	0	0	1 HLA-C major ↓	0	0	0
	12 months (21 samples)	VOC UK	0	0	0	0	0	0
		VOC SA	0	0	0	0	0	0
		VOC BR	0	0	0	0	0	0
		Iceland	0	0	0	0	0	8 HLA-A =2 1 HLA-DR mild ↓
Iran	6 months (23 Samples)	VOC UK	0	0	0	1 HLA-DR mild ↓	0	0
		VOC SA	0	0	0	0	0	0
		VOC BR	0	0	1 HLA-C major ↓	0	0	0
		Iran	1 HLA-C mild ↓	0	1 HLA-A mild ↓	1 HLA-DQ mild ↓	0	1 HLA-DR mild ↓
	12 months (12 samples)	VOC UK	0	0	0	0	0	0
		VOC SA	0	0	0	0	0	0
		VOC BR	0	0	0	0	0	0
		Iran	0	0	0	0	3 HLA-DQ mild ↓	0
Japan	6 months (22 samples)	VOC UK	0	0	0	1 HLA-DR mild ↓	0	0
		VOC SA	0	0	0	0	0	0
		VOC BR	0	0	0	0	0	0
		Japan	0	0	0	0	0	1 HLA-DR mild ↓
	12 months (22 samples)	VOC UK	0	0	0	0	0	0
		VOC SA	0	0	0	0	0	0
		VOC BR	0	0	0	0	0	0
		Japan	0	0	0	1 HLA-DR modest↓	0	6 HLA-DQ modest ↓

Summary Statement on Mutations related to Immunity: Overall, the mutations in the Iranian sequences were less predictable than those occurring in the other 2 countries based on NS/S ratios and C-> T(U) ratios. The C->T (U) mutations appeared to accrue over time (28.8%, 29.5%, and 46.6% vs 43.5%, 39.3% and 47.7% of total NT mutations for Iceland, Iran, and Japan), indicating that mutations that reduce the ability of the innate immune system to respond to viral RNA remained a major factor over time for all 3 populations. At both time points, mutations in antibody immunodominant epitopes were significantly more prevalent in sequences from Iran compared to those sequences from Iceland and Japan. Mutations that reduced the affinity of immunodominant CD4+ T helper epitopes (in S and N proteins) became significantly enriched in Japan and Iceland at the second time point, indicating that over time, the circulating strains in these 2 closed populations were adapting against the highest frequency HLA-alleles. This is important as CD4+ T cells control the B cell (antibody) response and CD8+ T cell cytotoxic response. In Iran, we observed a reoccurring (3/12) M protein mutation that mildly reduced binding affinity to HLA-II. **This M protein mutation was intriguing to us in that M protein is a relatively conserved protein compared to S and N and yet none of the mutations affected immunodominant epitopes defined for S and N.** Looking more specifically at S protein, which is the target structure used current vaccine strategies and antibody prophylaxis, the mutation frequency at the NT was similar across all 3 regions, though at both time points, many more mutations in the Spike protein were occurring in antibody binding sites in sequences from Iran compared to Japan and Iceland. **This implies that the selection force driving** the changes to the Iranian mutant sequences is particularly focused on reducing efficacy to neutralizing antibody response from vaccine measures whereas most of the mutations occurring Iceland and Japan appear to minimally influence immunity conferred by the vaccines in current use.

1. GISAID EpiCoV™ Database
2. NCBI GenBank
3. Lu S, Xie XX, Zhao L, et al. The immunodominant and neutralization linear epitopes for SARS-CoV-2. *Cell Rep*. 2021;34(4):108666.
4. Poh, C.M., Carissimo, G., Wang, B. *et al.* Two linear epitopes on the SARS-CoV-2 spike protein that elicit neutralising antibodies in COVID-19 patients. *Nat Commun* **11,** 2806 (2020).
5. Shrock E, Fujimura E, Kula T, et al. Viral epitope profiling of COVID-19 patients reveals cross-reactivity and correlates of severity. *Science*. 2020;370(6520):eabd4250.
6. Tarke A, Sidney J, Kidd CK, et al. Comprehensive analysis of T cell immunodominance and immunoprevalence of SARS-CoV-2 epitopes in COVID-19 cases. Preprint. *bioRxiv*. 2020;2020.12.08.416750.