Received: 10 December 2021 | Revised: 25 March 2022 | Accepted: 29 April 2022

DOI: 10.1111/zph.12971

ORIGINAL ARTICLE

WILEY

Geographical distribution of SARS-CoV-2 amino acids mutations and the concomitant evolution of seven distinct clades in non-human hosts

Murali Aadhitya Magateshvaren Saras | L. Ponoop Prasad Patro | Patil Pranita Uttamrao | Thenmalarchelvi Rathinavelan

Department of Biotechnology, Indian Institute of Technology Hyderabad, Kandi, India

Correspondence
Thenmalarchelvi Rathinavelan, Department of Biotechnology, Indian Institute of Technology Hyderabad, Kandi, Telangana State 502284, India.
Email: tr@bt.iith.ac.in

Funding information The authors thank BIRAC-SRISTI GYTI (PMU_2017_010) and BIRAC-SRISTI (PMU2019/007) for financial support. MAMS and LPPP thank MHRD for fellowship. PPU thank CSIR for fellowship.

Abstract

Since its first emergence in December 2019, the world has witnessed the eruption of mutations in the SARS-CoV-2 genome that have led to increased viral transmissibility and pathogenicity due to sustained local viral transmission. Zooanthroponotic and zoonotic transmissions have further raised concerns as they could result in the emergence of viral variants with a novel antigenicity and transmissibility that could jeopardize the vaccine efficacy. To understand the viral evolution during such transmissions, 1016 whole-genome sequences (deposited in GISAID as of March 7, 2022) (from 18 countries) corresponding to mink, cat, deer, dog, hyena, tiger, lion, gorilla, Syrian hamster, leopard cat, fishing cat, bear cat, coati, ferret, snow leopard and green monkey have been analysed here. Intriguingly, phyloproteome analysis indicate that Nsp2:R218C, Nsp2:D268-(deletion), Spike:D614G, Nsp12:P323L, Nsp2:A192V, ORF3a protein:Q57H, N protein:R203K and N protein:G204R/L, Spike:A222V, ORF10 protein:V30L and N protein:A220V are moderate or high recurring and clade decisive mutations, leading to 6 primary clades during the early stage of pandemic. Most interestingly, the human evolved delta variant having a combination of 26 (clade decisive) mutations defines the seventh clade and transmits to non-human hosts across the globe without exhibiting any country-specific mutation(s). Nonetheless, Spike:D614G and Nsp12:P323L together with (i)N protein:R203K,N protein:G204R/L,Spike:V70-, Spike:H69-, Nsp12:T739I, and Nsp1:M85-, (ii)Nsp2:A192V, Nsp3:D178Y, (iii)Nsp2:T85I, N protein:P67S and ORF3a protein:Q57H and (iv)Spike:A222V, ORF10 protein:V30L, N protein:A220V and Spike:F486I are specific to Denmark, Netherlands, USA and Latvia respectively and, (v)Nsp2:D268- and Nsp13:R292C that are devoid of Spike:D614G and Nsp12:P323L is specific to Netherlands. SARS-CoV-2 variants consisting of these mutations are also seen in the human SARS-CoV-2 sequences from the same country. Independent country-specific SARS-CoV-2 variant evolution further indicates distinct epidemiological dynamics during zooanthroponotic and zoonotic transmissions. Thus, the results presented here indicate the need for the surveillance of viral evolution in non-human hosts also during the future pandemic.

L. Ponoop Prasad Patro and Patil Pranita Uttamrao contributed equally.

KEYWORDS
Non-human hosts, Proteome evolution, SARS-CoV-2

1 | INTRODUCTION

The evidences about the ability of SARS-CoV-2 to transmit from animal to human (Lytras et al., 2021) and vice-a-versa (Hoffmann et al., 2021) are alarming as the virus can acquire new mutations to adapt to a new host. Such acquired mutations can trigger an outbreak in animal or human population when the favourable condition for the dissemination is met. In fact, several SARS-CoV-2 epizootics and the associated mutations in mink have been reported during this ongoing pandemic (Devaux et al., 2021, Eckstrand et al., 2021). Such mutations may make the virus more contagious, virulent and infectious that could further jeopardize the vaccine efficacy.

Until now, mink (Oude Munnink et al., 2021), lion (McAloose et al., 2020), tiger (McAloose et al., 2020), cat (Sailleau et al., 2020), dog (Sit et al., 2020), Syrian hamster (Hui-Ling Yen et al., 2022, Haagmans, 2022) and ferret (Kim et al., 2020) have been reported to be the non-human hosts for SARS-CoV-2 (Sharun et al., 2021, Decaro et al., 2021, Hamer et al., 2021). Since the origin of pandemic is believed to be zoonotic (Lytras et al., 2021), independent evolution of SARS-CoV-2 mutation in minks in Denmark farms and the subsequent zoonotic transfer raises concerns (Oude Munnink et al., 2021, Sharun et al., 2021, Larsen & Paludan, 2020). It has also been reported that the virus evolves at a faster rate in mink than in humans (Oude Munnink et al., 2021). This indeed has led to the mass culling of 17 million minks in Denmark (Frutos & Devaux, 2020).

Many studies have shown that many wild and domesticated species can become susceptible to SARS-CoV-2 (Goraichuk et al., 2021, Liu et al., 2020, Luan et al., 2020, Shi et al., 2020). In fact, a recent study has suggested that ACE2 (receptor for viral entry) orthologs in 252 mammals and 72 birds can facilitate its entry into those species (Samavati & Uhal, 2020). Such transmissions among different hosts (both non-human and human) may facilitate the virus to evolve with fitness-enhancing mutations and pave the way for the emergence of new variants. Such variants may have more pathogenicity, transmissibility and immune evading capacity, thus may reduce vaccine efficacy. Under such a scenario, the non-human hosts can become a viral reservoir and may develop the potential for future zoonotic transfer(s).

In the perspective of understanding the viral evolution in the non-human hosts, 1016 SARS-CoV-2 whole-genome sequences (deposited in GISAID as on March 7, 2022) from 16 non-human hosts (mink (*Neovison vison* and *Mustela lutreola*), cat (*Felis catus*), dog (*Canis lupus familiaris*), fishing fat (Prionailurus-viverrinus), bear cat (Arctictis-binturong), hyena (Crocuta-crocuta), tiger (*Panthera tigris*, *Panthera tigris jacksoni* and *Panthera tigris sondaica*), lion (*Panthera_leo*), deer (*Odocoileus virginianus*), gorilla (Gorilla_gorilla_gorilla), green monkey (*Cercopithecus sabaeus)*, Syrian hamster (*Mesocricetus auratus*), leopard cat (*Prionailurus bengalensis euptilurus*), ferret (*Mustela putorius furo*), coati (*Nasua-nasua*) and snow leopard (*Panthera uncia*) collected from 18 countries have been analysed here. Although there are several investigations on the variations in mink SARS-CoV-2 sequences (Devaux et al., 2021, Oude Munnink et al., 2021), these analyses are confined either to a smaller fraction of sequences (van Dorp et al., 2020) or to a specific country (Eckstrand et al., 2021).

Impacts

- Phyloproteomic analysis of non-human host SARS-CoV-2 proteome sequences have led to 7 major clades.
- SARS-CoV-2 infects a broad range of hosts which enhances the chances of future spillover and spillback events.
- 10 and 44 protein variants are found to be transmitted through zoonotic and zooanthroponotic mode respectively which necessitates the tracking of mutations in all the extended hosts of SARS-CoV-2 to prevent the emergence of novel variants and cross-species transmission.

2 | MATERIALS AND METHODS

2.1 | Creation of a local database of non-human hosts SARS-CoV-2 pan proteome

A total of 1284 SARS-CoV-2 whole-genome sequences (which do not have undefined nucleotides ['N'] in the coding region) from 17 non-human hosts from 18 countries which are deposited in GISAID on or before 7 March 2022 were used for the analysis. Following options of GISAID database were used in downloading the sequences: (i) complete sequences, (ii) high coverage, (iii) have complete collection date information and iv) sequences belonging to induced SARS-CoV-2 infection were excluded (specifically mice). Figure S1 shows the organism wise and country wise statistics of SARS-CoV-2 sequences used in the current investigation. Subsequently, the coding regions of 1016 sequences from 16 non-human hosts were translated into individual proteins using the in-house scripts and a local database of non-human hosts SARS-CoV-2 pan proteome was created (Table S1). Note that sequences with insertions were not considered (Table S2).

Ethical approval is not required as the sequences analysed in the study were collected from publicly available database, and the methodology approach was entirely computational.

2.2 | Mutation and variant analyses of non-human hosts SARS-CoV-2 proteome sequences

By considering the translated proteome sequence of the first published SARS-CoV-2 genome sequence (Genbank ID: NC_045512.2) as the reference sequence, the sequences in the local database were subjected to the mutational occurrence analysis. Following the definition used in the previous investigations, high (HR), moderate (MR) and low (LR) recurring mutations (Patro et al., 2021a, Patro et al., 2021b, Uttamrao et al., 2021) were calculated based on the percentage frequency (PF) of occurrence in 1016 sequences. A mutation was defined as high, moderate and low recurring respectively if it occurred above 10%, 1%–10% and <1% (but, occurring at least 3 times) percentage frequency. Pymol software (The PyMOL Molecular Graphics System, Version 2.0, Schrödinger, LLC) was used to project the HR mutations onto the 3D structures of SARS-CoV-2 proteins (Figure S2) that were either I-TASSER generated (https://zhanggroup.org//COVID-19/) or deposited in PDB. Subsequently, non-human hosts SARS-CoV-2 variant (combination of different amino acids mutations in the proteome (Patro et al., 2021a) analysis was performed. Further, frequency of different unique variants was also calculated.

2.3 | Construction of non-human SARS-CoV-2 phyloproteomic tree

Using the data obtained from the variant analysis, 546 unique variants (non-redundant) were identified and subjected to non-human SARS-CoV-2 phyloproteomic analysis. Subsequently, the non-redundant whole proteomes were subjected to alignment using MAFFT (Katoh & Standley, 2013). The aligned sequences were then used in the construction of the country-based and organism-based phyloproteomic trees using IQ-TREE software (Nguyen et al., 2015) by employing the maximum likelihood method. The phylorproteogram was visualized and analysed using iTOL tool (Letunic & Bork, 2019).

The graphs used in this study were created using Microsoft Excel 16. Inkscape software was used for organizing the figures.

3 | RESULTS

To track the viral evolution in 16 non-human hosts (mink, cat, deer, dog, Hyena, tiger, lion, gorilla, green monkey, Syrian hamster, leopard cat, fishing cat, bear cat, coati, ferret and snow leopard), the whole proteome analysis of 1016 SARS-CoV-2 sequences has been carried out to identify the high and moderate recurring mutations, presuming that they have a positive selection pressure.

3.1 | Evolution of non-human hosts SARS-CoV-2 clades

A total of 48 high (Table 1, Figure S2), 156 moderate and 208 low recurring mutations are identified in 1016 SARS-CoV-2 sequences corresponding to 16 non-human hosts considered in this investigation (Table S3). Not surprisingly, Spike:D614G and Nsp12:P323L are found with a percentage frequency of 95.82%. There is a major diversion in the phyloproteogram because of the sequences that are devoid of this pair of mutations (clade nodes 1 and 2 in Figures 1 and 2) and the ones that have them (clade node 3 in Figures 1 and 2). Due to the highest percentage frequency, the clade (node 3) having Spike:D614G and Nsp12:P323L has emerged as the major one. Among the HR and MR mutations, 4 HR (Nsp2:A192V (node 5), ORF3a protein:Q57H (node 6) and N protein:R203K and N protein:G204R (node 8)) and 4 MR (ORF3a protein:Q57K (part of node 4) and, Spike:A222V, ORF10 protein:V30L and N protein:A220V (node 9)) mutations define 5 distinct diversions (nodes 4, 5, 6, 8 and 9 in Figures 1 and 2) under the clade defined by Spike:D614G and Nsp12:P323L (node 3). In addition to that, simultaneous occurrence of 26 HR mutation leads to node 7, which is defined as delta variant by the WHO. Node 5 diverges into node 10 due to the moderately recurring Nsp3:D178Y mutation. Node 6 further diverges into two nodes due to the highly recurring T85I (node 11) and moderately recurring Nsp6:A46V, Nsp2:R218C and Nsp9:G37R (node 12) mutations. Similarly, node 8 and node 9 diverge due to the highly recurring Spike:H69- & Spike:V70- (node 13) and moderately recurring Spike:F486I (node 14) mutations respectively. Divergence of node 10 into node 15 (Spike:A262S and Nsp13:I258V), node 11 into node 16 (N protein:P67S) and node 13 into node 17 (defined by Nsp12:T739I and Nsp1:M85-) are seen due to the moderate or high recurring mutations. Additional divergence of node 17 into nodes 18 (Nsp2:P181S) and 19 (Nsp3:L1244F) are also seen. Since the other divergences in the phyloproteogram occur with a percentage frequency lesser than 4, they are not considered for the clade node definition, thus, are not discussed here. Exceptionally, node 2 is considered for the discussion (PF of decisive mutation = 3.5%) as it leads to a unique clade lacking Spike:D614G and Nsp12:P323L and has Netherlands specific Nsp13:R392C and Nsp2:D268- mutations.

Some of the clades corresponding to the above nodes are due to country-specific mutations as seen in humans (Patro et al., 2021a). For instance, N protein:R203K and N protein:G204R that are predominantly found in Denmark with the percentage frequency of 34.07% and 32.04% respectively have led to node 8. Further, a few divergences seen in this clade are due to the inclusion of additional mutations which are either predominantly (node 13) or specifically (nodes 17, 18 and 19) found in Denmark.

The next prominent branch in the phyloproteogram is due to the presence of ORF3a protein:Q57H (percentage frequency = 21.66%) (node 6) and Nsp2:A192V (PF = 11.49%) (node 5) which lacks N protein:R203K and N protein:G204R/L. ORF3a protein:Q57H (node 6) is predominant in USA and Netherlands, so as the concomitant branching. Node 12 emerging from the node 6 is specific to Netherlands. While the inclusion of Nsp2:T85I (node 11) to ORF3a protein:Q57H is specific to USA and Netherlands, further divergence (node 16) seen in the node 11 is specific to USA (N protein:P67S). Interestingly, Denmark SARS-CoV-2 mink sequences do not have Nsp2:A192V, ORF3a protein:Q57H, but have N protein:R203K and

TABLE 1 List of HR mutations found in the non-human hosts SARS-CoV-2 sequences analysed in the current investigation

	Mutation	PF (%)	Node number(s) (if it is a clade decisive mutation)	Top 5 lineages
1	Spike:D614G	95.82	3–19	B.1.1.298, B.1, B.1.8, B.1.177.60, B.1.22,B.1.2
2	Nsp12:P323L	95.72	3–19	B.1.1.298, B.1, B.1.8, B.1.177.60, B.1.22
3	ORF3a protein:H182Y	41.40	–	B.1.1.298, B.1.8, B.1, B.1.22, AU.1
4	Spike:Y453F	37.94	–	B.1.1.298, B.1, B.11, B.1.177.60, B.1.8
5	Nprotein:S194L	36.62	–	B.1.1.298, B.1, B.1.234, B.1.243, B.1.2
6	N protein:R203K	34.07	8,13,17–19	B.1.1.298, B.1.1.7, B.1.1, B.1.1.464, B.1.1.294
7	Nsp3:N1263-	33.57	–	B.1.1.298, B.1.8
8	N protein:G204R	32.04	8,13,17–19	B.1.1.298, B.1.1.7, B.1.1.294, Q.1, B.1.1.486
9	Spike:V70-	30.62	13,17–19	B.1.1.298, B.1.1.7, B.1.177.60, Q.1, B.1.258
10	Spike:H69-	30.62	13,17–19	B.1.1.298, B.1.1.7, B.1.177.60, Q.1, B.1.258
11	Nsp12:T739I	28.38	17–19	B.1.1.298
12	Nsp1_:M85-	27.77	17–19	B.1.1.298, B.1.240, B.1
13	ORF3a protein:Q57H	21.66	6, 11, 12, 16	B.1, B.1.22, B.1.2, B.1.311, B.1.526
14	Spike:G142D	17.59	–	B.1, AY.44, AY.44, AY.3, AY.20
15	Nsp2:T85I	15.97	11, 16	B.1, B.1.2, B.1.311, B.1.526, B.1.582
16	N protein:D377Y*	15.86	7	AY.44, AY.103, B.1.2, AY.3, AY.20
17	Nsp3:L1244F	15.05	19	B.1.1.298
18	Nsp13:P77L*	14.44	7	AY.44, AY.103, AY.3, AY.20, AY.43
19	Nsp12:G671S*	14.44	7	AY.44, AY.103, AY.3, AY.20, AY.43
20	Mprotein:I82T*	14.44	7	AY.44, AY.103, AY.3, AY.20, AY.43
21	Spike:T478K*	14.34	7	AY.44, AY.103, AY.3, AY.20, AY.43
22	Spike:T19R*	14.34	7	AY.44, AY.103, AY.3, AY.20, AY.43
23	Spike:L452R*	14.34	7	AY.44, AY.103, AY.3, AY.20, AY.43
24	ORF3a protein:S26L*	14.34	7	AY.44, AY.103, AY.3, AY.20, AY.43
25	Nprotein:R203M*	14.24	7	AY.44, AY.103, AY.3, AY.20, AY.43
26	Spike:P681R*	14.14	7	AY.44, AY.103, AY.3, AY.20, AY.43
27	ORF3a protein:L219V	14.14	-	B.1.177.60, B.1, B.1.22, B.1.8, B.1.160
28	Spike:R158-*	14.03	7	AY.44, AY.103, AY.3, AY.20, AY.43
29	Spike:F157-*	14.03	7	AY.44, AY.103, AY.3, AY.20, AY.43
30	Spike:E156G*	14.03	7	AY.44, AY.103, AY.3, AY.20, AY.43
31	Spike:D950N*	14.03	7	AY.44, AY.103, AY.3, AY.20, AY.43
32	Nsp3:P1228L*	13.93	7	AY.44, AY.103, AY.3, AY.20, AY.43
33	ORF7a protein:T120I*	13.53	7	AY.44, AY.103, AY.3, AY.20, AY.43
34	Nsp3:A488S*	13.53	7	AY.44, AY.103, AY.3, AY.20, AY.43
35	ORF7a protein:V82A*	13.42	7	AY.44, AY.103, AY.3, AY.20, AY.43
36	Nsp4:V167L*	13.42	7	AY.44, AY.103, AY.3, AY.20, AY.43
37	Nsp4:T492I*	13.42	7	AY.44, AY.103, AY.3, AY.20, AY.43
38	Nsp14:A394V*	13.42	7	AY.44, AY.103, AY.3, AY.20, AY.43
39	N protein:D63G	13.42	7	AY.44, AY.103, AY.3, AY.20, AY.43
40	ORF7b protein:T40I	13.32	7	AY.44, AY.103, AY.3, AY.20, AY.43
41	Nsp6:T77A	13.32	7	AY.44, AY.103, AY.3, AY.20, AY.43
42	Nsp3:P1469S	13.32	7	AY.44, AY.103, AY.3, AY.20, AY.43
43	N protein:G215C	13.32	7	AY.44, AY.103, AY.3, AY.20, AY.43
44	Spike:F486L	11.80	–	B.1.8, B.1, B.1.22, B.11
45	Nsp9:G37E	11.69	–	B.1.8, B.1, B.11, B.1.22, AY.4

TABLE 1 (Continued)

	Mutation	PF (%)	Node number(s) (if it is a clade decisive mutation)	Top 5 lineages
46	Nsp2:A192V	11.49	5, 10, 15	B.1.8, B.1.243
47	Spike:N501T	11.29	–	B.1, B.1.177.60, B.1.536, B.1.8, B.1.22
48	Nsp6:L37F	10.27	–	B.1.1.298, B.1.8, B.1.22, B.1.234, B

Note: * indicates the delta variant (WHO nomenclature).

N protein:G204R/L, whereas the vice-a-versa is seen in Netherlands mink sequences (Figure S3). Similarly, branches 10 and 15 are specific to Netherlands. Since node 9 has resulted out of mutations that are majorly seen in Latvia (among which, Spike:A222V, ORF10 protein:V30L and N protein:A220V are reported to be co-occurring in human[21]), further divergence in this node (node 14) is also specific to Latvia.

A small population of sequences in node 4 lack any high or moderate recurring mutations (other than Spike:D614G and Nsp12:P323L) which is seen in different organisms. Four sequences from USA that do not have any of the aforementioned mutations stay close to the reference sequence (Genbank ID: NC_045512.2) (node 1). Thus, the phyloproteomic analysis of non-human host's SARS-CoV-2 sequences indicates that 9 out of 19 clade nodes are due to the country-specific clade definitive mutations (nodes 2, 10, 12, 14–19) (Figures 1, 2) and are found during pre-delta variant era. However, such country-specific mutations are absent in the delta variant (node 7).

3.2 | Comparative analysis of SARS-CoV-2 variant distribution in human and non-human hosts

To understand the compatibility of SARS-CoV-2 variants (sequences having combination of LR, HR and MR) in human and non-human hosts, the presence of non-human SARS-CoV-2 variants in human SARS-CoV-2 sequences have been analysed. There are 546 unique variants identified in SARS-CoV-2 sequences present in non-human hosts (Table S4), wherein, 45 is the highest number of mutations seen in a variant (GISAID: EPI_ISL_9347717 & EPI_ISL_9347721). One among the unique variants which has 13 highly recurring mutations (Nsp1:M85-; Nsp3:L1244F; Nsp3:N1263-; Nsp12:P323L; Nsp12:T739I; Spike:H69-; Spike:V70-; Spike:Y453F; Spike:D614G; ORF3a protein:H182Y; N protein:S194L; N protein:R203K; N protein:G204R) occurs 65 times in Denmark mink sequences during September 2020 and November 2020. Similarly, 20 and 19 are the highest occurrence of a variant in USA (Nsp2:T85I; Nsp2:V381A; Nsp6:Q208R; Nsp9:T24I; Nsp9:T34I; Nsp12:P323L; Nsp16:T140I; Spike:G142D; Spike:Y145-;Spike:F486L; Spike:N501T; Spike:D614G; ORF3aprotein:Q57H; ORF3aprotein:L219V; Eprotein:C44Y; Nprotein:D103Y; Nprotein:R191L) and Netherlands (Nsp2:A192V; Nsp3:D178Y; Nsp3:N1263-; Nsp9:G37E; Nsp12:P323L; Nsp13:I258V; Spike:A262S; Spike:Q314K; Spike:F486L; Spike:D614G; ORF3a protein:H182Y) respectively. Further, some of the variants in the same country are found to be common between different organisms, indicating their compatibility across different hosts.

Interestingly, analysis of the occurrence of non-human SARS-CoV-2 variants in human provides an interesting clue about both zooanthroponotic and zoonotic transmissions (Table S5). The hunt for the presence of non-human hosts SARS-CoV-2 variants in human has been carried out by considering 2,943,976 whole-genome sequences (deposited in GISAID as on 14 February 2022). The results indicate that there are 76 non-human SARS-CoV-2 variants are found to have at least a single occurrence in human. Importantly, the SARS-CoV-2 variant (having 13 highly recurring mutations) that has emerged in mink (GISAID: EPI_ISL_641477 deposited on 10/09/20) with 65 occurrences is also found to have 1 occurrence in human. This indicates the possible zoonotic transmission of this variant from mink to humans. Similarly, 44 variants first reported in human are also found to be transmitting to non-human hosts (Table S5). For instance, a variant (having Nsp2:T85I; Nsp12:P323L; Spike:D614G and ORF3a protein:Q57H mutations and, reported first in France on 01/01/20 GISAID: EPI_ISL_4537783) transmitted from human to tiger (reported first in France on 02/04/2020 GISAID: EPI_ISL_566043) is also found. Strikingly, this variant occurs 4948 times in human across 65 countries. Most importantly, among the different SARS-CoV-2 variants, delta variant is predominantly found to take part in zooanthroponotic transmission (Tables S5).

4 | DISCUSSION

The variations in SARS-CoV-2 genome since its first outbreak in Wuhan, China in 2019 and the consequent impact on the increased viral fitness (Plante et al., 2021, Yang et al., 2021), contagiousness (Chen et al., 2020, Zhang et al., 2020) pathogenicity (Bakhshandeh et al., 2021, Volz et al., 2021) and ability to escape from neutralizing antibody (Harvey et al., 2021, Mlcochova et al., 2021, Lazarevic et al., 2021, Garcia-Beltran et al., 2021) raise concerns. Additionally, the ability of SARS-CoV-2 to affect a broad range of hosts, the emergence of variants with fitness-enhancing mutations in the non-human hosts and the consequent zoonotic transmission are alarming (Fenollar et al., 2021) (https://www.who.int/news-room/q-a-detail/sars-cov-2-evolution). Indeed, the SARS-CoV-2 pandemic currently being faced by Hong Kong is due to the spread of delta variant (AY.127) from hamsters to human (Bart L Haagmans, 2022, Hui-Ling Yen et al., 2022). This necessitates the understanding of the viral evolution in the non-human hosts. Very recently, it has been shown

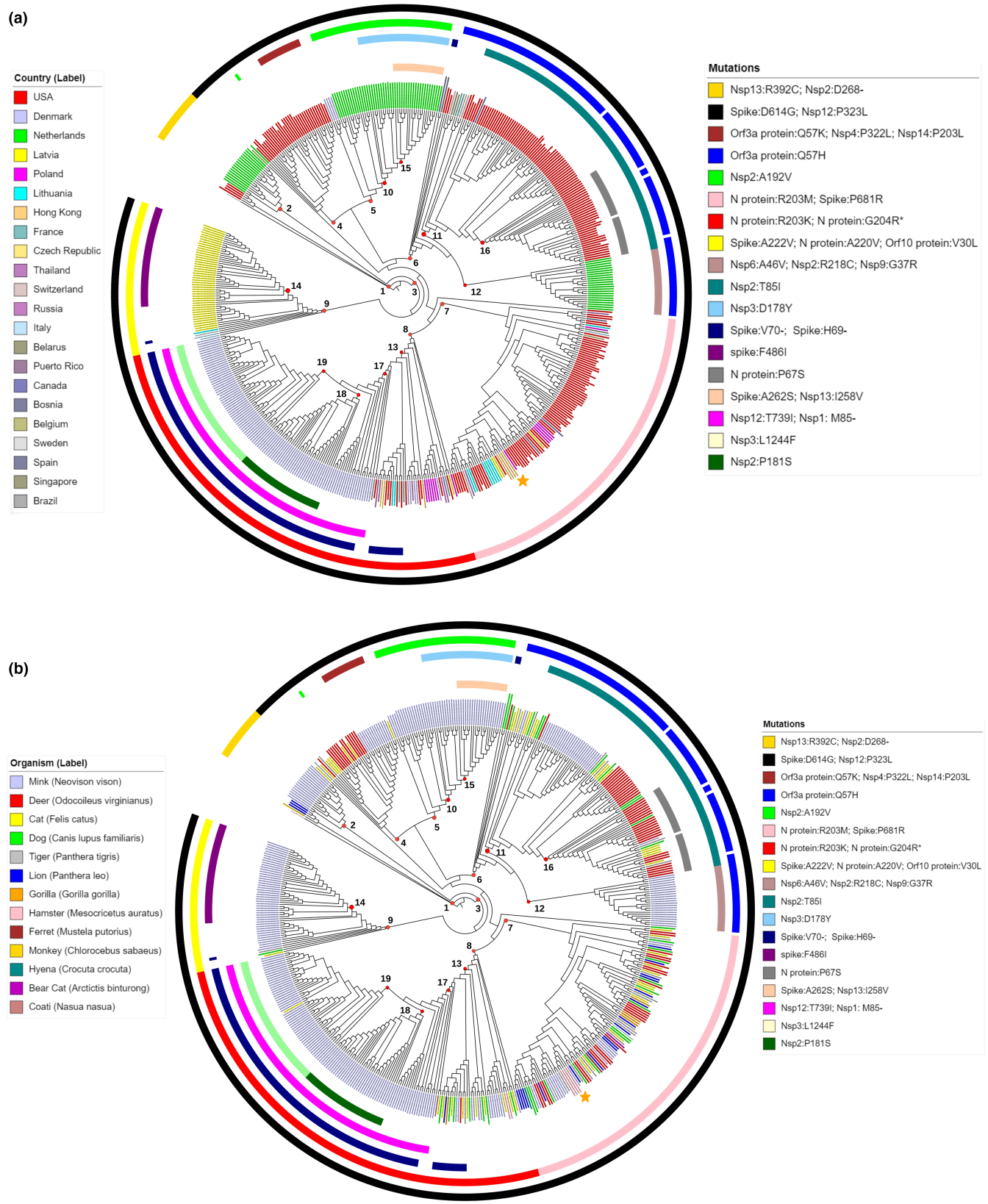

FIGURE 1 Phyloproteogram of SARS-CoV-2 sequences from non-human hosts coloured coded (left side) by (a) country wise and (b) organism wise. The numerical values (1–19) correspond to the nodes associated with the corresponding branches. Note that the colour coding on the right-hand side correspond to the decisive mutations (as they decide the divergence) which are indicated by the circles surrounding the phyloproteogram. For the sake of clarity, only N protein:R203M is alone indicated as a representative mutation of delta variant, and the other 25 HR mutations are not marked here. Refer Table 1 for the list of HR mutations. Note that Syrian hamster sequences that is responsible for the ongoing SARS-CoV-2 outbreak in Hong Kong falls under the node 7 (marked as a star)

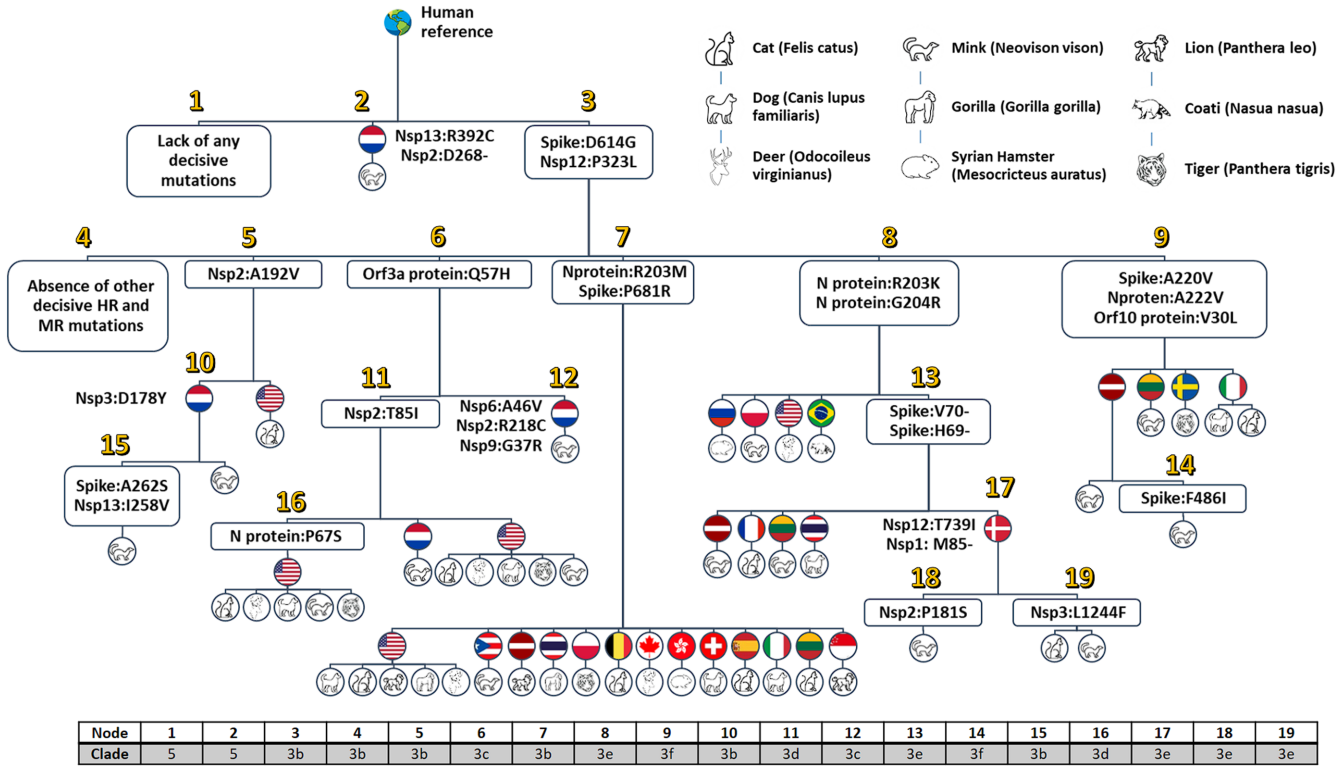

FIGURE 2 Flowchart depicting the branching of SARS-CoV-2 clades/sub-clades in non-human hosts phyloproteogram. Comparison of the non-human hosts SARS-CoV-2 clades (nodes 1–19) with the previously defined SARS-CoV-2 human clades (grey coloured) (Patro et al., 2021a) are given at the bottom. Note that the mutations specific to different countries are represented by their flags and the corresponding non-human hosts are indicated pictorially

from this laboratory that there are important recurring amino acid mutations in the human SARS-CoV-2 proteome which have led to significant divergences in SARS-CoV-2 proteome (Patro et al., 2021a, Patro et al., 2021b). The same methodology has been adopted here to identify the recurring mutations in non-human SARS-CoV-2 proteome and their role in representing the divergence in the phyloproteome of SARS-CoV-2 by considering the viral sequences isolated from mink, cat, deer, dog, Hyena, tiger, lion, gorilla, green monkey, Syrian hamster, leopard cat, fishing cat, bear cat, coati, ferret and snow leopard in 18 countries (Figure 1). It is noteworthy that among the 1016 whole-genome sequences consider here for the analysis, ~70% are from the mink host alone.

The analysis indicates a total of 48 (Figure S2), 156 and 208 high, moderate and low recurring mutations (Table S1) in the SARS-CoV-2 sequences obtained from 16 non-human hosts. The results also indicate that 192 recurrent (LR, MR and HR) mutations are found to be host specific, most of which belongs to mink and deer followed by Syrian hamster (Figure 2, Table S6). Among them, 1 HR (Nsp9:G37E in mink) and 87 MR are found to be host specific. Intriguingly, out of the 48 HR and 156 MR mutations, only 34 (Spike:D614G, Nsp12:P323L, NSp2:T85I, N protein:R203K, N protein:G204R, Nsp2:A192V, ORF3a protein:Q57H, Spike:H69-, Spike:V70-, Nsp12:T739I, Nsp1:M85-, Nsp3:L1244F and 26 simultaneously occurring delta variant (node 7) mutations (Refer Table 1)) HR and 14 (Nsp3:D178Y, Spike:A262S, Nsp13:I258V, Nsp6:A46V, Nsp12:R218C, Nsp9:G37R, N protein:P67S, Nsp2:P181S, N protein:A220V, Spike:A222V, ORF10 protein:V30L, Spike:F486I, Nsp13:R392C and Nsp2:D268-) MR mutations act as the decisive mutations for the divergence (viz., nodes 1–19) in the phyloproteogram (Figure 2). Syrian hamster sequences that is responsible for the ongoing SARS-CoV-2 outbreak in Hong Kong falls under the node 7. Y453F which has been identified to be evolved in mink and transmitted to human (Mallapaty, 2020) is not a decisive mutation (despite being a HR with the percentage frequency of 37.94%) as it is distributed across different clades. Similarly, Spike:N501Y (alpha variant) that has emerged in human (Leung et al., 2021) is also seen with moderate recurrence in mink (together in deer, cat, dog, lion, tiger and gorilla) but does not act as a decisive mutation. Spike:A222V, ORF10 protein:V30L and N protein:A220V mutations representing 20A-EU1 clade (clade 3f of human, node 9 of Figures 1 and 2) has further diverged in mink with the occurrence of Spike:L18F, Nsp12:G823F and ORF3a protein:V77F mutations upon zooanthroponotic transmission (Figures 1 and 2). Yet another point to be noted is the presence of human SARS-CoV-2 clades 3b-3f (Figure 2) (Patro et al., 2021a) in the non-human hosts indicating the positive selection pressure of the decisive mutations in both the human and non-human hosts. Importantly, further divergence in SARS-CoV-2 clades (for instance, nodes 3 (clade 3a of humans), 6 (clade 3c of humans) and 8 (clade 3e of humans)) after the zooanthroponotic transmission and their high or moderate recurrence are also found (Figure 2).

Strikingly, some of the mutations are found to be country-specific resulting in region specific evolution of SARS-CoV-2 in non-human hosts during the pre-delta variant era (Figures 1, 2). However, the delta variant (node 7) does not have such country-specific mutations and is found to be present in several non-human hosts across different countries (Figure 2). The smart choice of a combination of 26 clade decisive mutations by the virus in the delta variant perhaps be the reason for such a broad range of hosts.

The analysis further indicates the presence of non-human hosts SARS-CoV-2 variants in human. It is also revealed from the current study that 76 non-human SARS-CoV-2 variants are found in human, wherein 10 of them are transmitted through zoonotic transmission (Table S5). Alarmingly, one of the variants that has emerged in mink (GISAID: EPI_ISL_641477 deposited on 10/09/20) with 65 occurrences is also found to occur once in the humans (month of first occurrence: Nov-2020, GISAID: EPI_ISL_6236282).

Thus, the results presented here indicate the possibility of zoonotic and zooanthroponotic transmissions and the concomitant emergence of HR and MR mutations (either host-specific or non-host-specific). This is suggestive of evolution of fitness-enhancing variants (viz., mutation combinations) in non-human hosts in the future which can be transmitted to human through zoonotic spillover as recently witnessed in Hong Kong (Haagmans, 2022, Hui-Ling Yen et al., 2022). Thus, the non-human hosts can act as a reservoir for future SARS-CoV-2 zoonotic transmissions and necessitates the need to track the SARS-CoV-2 variant evolution in non-human host and speed-up the non-human host SARS-CoV-2 vaccine development program.

5 | CONCLUSIONS

The variations in non-human hosts (mink, cat, deer, dog, Hyena, tiger, lion, gorilla, green monkey, Syrian hamster, leopard cat, fishing cat, bear cat, coati, ferret and snow leopard) SARS-CoV-2 pan proteome (18 countries) have been investigated here to track the viral evolution in non-human hosts. Such investigations are important to understand the role of mutations in viral fitness, contagiousness and pathogenicity in non-human hosts. Intriguingly, the results show seven major divergence in the non-human hosts SARS-CoV-2 phyloproteome due to 30 (among which, 26 are specific to delta and most of the time occur simultaneously) high and 4 moderate recurring clade decisive mutations. Some of the mutations are specific to certain countries during the pre-delta variant era, thus indicating the possibility of country-specific epidemiological dynamics SARS-CoV-2 in non-human hosts. Further, the clades that are seen to be favourable in humans are also seen in non-human hosts indicative of the positive selection pressure of these mutations in a broad range of hosts. Some of the non-human hosts clades have indeed diverged from the corresponding human clades due to the inclusion of HR or MR mutations. Considering the zoonotic origin of the ongoing pandemic along with the witnessed SARS-CoV-2 zooanthroponotic, zoonotic and epizootic transmissions, the results presented here indicate the possibility of evolution of new variants that are compatible in both human and non-human hosts and can deceive vaccines. Thus, the results presented here indicate the necessity of tracking the viral evolution not only in the human host but also in the non-human hosts during the future pandemic as the zoonotic and zooanthroponotic transmission may lead to independent viral evolution.

AUTHOR CONTRIBUTIONS

MAMS has collected, analysed and plotted the data. LPPP tweaked the codes and PPU rendered support for phyloproteome analysis. TR and PPU wrote the manuscript. TR conceptualized and supervised the project.

ACKNOWLEDGEMENTS

The authors acknowledge all the researchers who have deposited the SARS-CoV-2 genome sequences (used in this study) to GISAID and GISAID for providing the sequences (Table S7). IIT Hyderabad and CDAC-India are acknowledged for the computational resources.

CONFLICT OF INTEREST

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

DATA AVAILABILITY STATEMENT

The data that supports the findings of this study are available in the supplementary material of this article

REFERENCES

Bakhshandeh, B., Jahanafrooz, Z., Abbasi, A., Goli, M. B., Sadeghi, M., Mottaqi, M. S., & Zamani, M. (2021). Mutations in SARS-CoV-2; Consequences in structure, function, and pathogenicity of the virus. *Microbial Pathogenesis*, *154*, 104831.

Chen, J., Wang, R., Wang, M., & Wei, G. W. (2020). Mutations strengthened SARS-CoV-2 infectivity. *Journal of Molecular Biology*, *432*, 5212–5226.

Decaro, N., Vaccari, G., Lorusso, A., Lorusso, E., De Sabato, L., Patterson, E. I., Di Bartolo, I., Hughes, G. L., Teodori, L., Desario, C., Colitti, B., Ricci, D., Buonavoglia, D., Rosati, S., Martella, V., Camma, C., Agrimi, U., & Elia, G. (2021). Possible human-to-dog transmission of SARS-CoV-2, Italy, 2020. *Emerging Infectious Diseases*, *27*, 1981–1984.

Devaux, C. A., Pinault, L., Delerce, J., Raoult, D., Levasseur, A., & Frutos, R. (2021). Spread of mink SARS-CoV-2 variants in humans: A model of sarbecovirus interspecies evolution. *Frontiers in Microbiology*, *12*, 675528.

van Dorp, L., Tan, C. C. S., Lam, S. D., Richard, D., Owen, C., Berchtold, D., Orengo, C., & Balloux, F. (2020). Recurrent mutations in SARS-CoV-2 genomes isolated from mink point to rapid host-adaptation. *bioRxiv*. https://doi.org/10.1101/2020.11.16.384743

Eckstrand, C. D., Baldwin, T. J., Rood, K. A., Clayton, M. J., Lott, J. K., Wolking, R. M., Bradway, D. S., & Baszler, T. (2021). An outbreak of SARS-CoV-2 with high mortality in mink (Neovison vison) on multiple Utah farms. *PLoS Pathogens*, *17*(11), e1009952.

Fenollar, F., Mediannikov, O., Maurin, M., Devaux, C., Colson, P., Levasseur, A., Fournier, P. E., & Raoult, D. (2021). Mink, SARS-CoV-2, and the human-animal interface. *Frontiers in Microbiology*, *12*, 663815.

Frutos, R., & Devaux, C. A. (2020). Mass culling of minks to protect the COVID-19 vaccines: is it rational? *New Microbes and New Infections*, *38*, 100816.

Garcia-Beltran, W. F., Lam, E. C., St Denis, K., Nitido, A. D., Garcia, Z. H., Hauser, B. M., Feldman, J., Pavlovic, M. N., Gregory, D. J., Poznansky, M. C., Sigal, A., Schmidt, A. G., Iafrate, A. J., Naranbhai, V., & Balazs, A. B. (2021). Multiple SARS-CoV-2 variants escape neutralization by vaccine-induced humoral immunity. *Cell*, *184*(9), 2372–2383.e9.

Goraichuk, I. V., Arefiev, V., Stegniy, B. T., & Gerilovych, A. P. (2021). Zoonotic and reverse zoonotic transmissibility of SARS-CoV-2. *Virus Research*, *302*, 198473.

Haagmans, B. L., & Koopmans, M. P. G. (2022). Spreading of SARS-CoV-2 from hamsters to humans. *The Lancet*, *399*, 1027–1028.

Hamer, S. A., Ghai, R. R., Zecca, I. B., Auckland, L. D., Roundy, C. M., Davila, E., Busselman, R. E., Tang, W., Pauvolid-Correa, A., Killian, M. L., Jenkins-Moore, M., Torchetti, M. K., Robbe Austerman, S., Lim, A., Akpalu, Y., Fischer, R. S. B., Barton Behravesh, C., & Hamer, G. L. (2022). SARS-CoV-2 B.1.1.7 variant of concern detected in a pet dog and cat after exposure to a person with COVID-19, USA. *Transboundary and Emerging Diseases*, *69*(3), 1656–1658.

Harvey, W. T., Carabelli, A. M., Jackson, B., Gupta, R. K., Thomson, E. C., Harrison, E. M., Ludden, C., Reeve, R., Rambaut, A., Consortium, C.-G. U., Peacock, S. J., & Robertson, D. L. (2021). SARS-CoV-2 variants, spike mutations and immune escape. *Nature Reviews Microbiology*, *19*, 409–424.

Hoffmann, M., Zhang, L., Kruger, N., Graichen, L., Kleine-Weber, H., Hofmann-Winkler, H., Kempf, A., Nessler, S., Riggert, J., Winkler, M. S., Schulz, S., Jack, H. M., & Pohlmann, S. (2021). SARS-CoV-2 mutations acquired in mink reduce antibody-mediated neutralization. *Cell Reports*, *35*, 109017.

Hui-Ling Yen, T. H. C. S., Brackman, C. J., Chuk, S. S. Y., Haogao, G., Tam, K. W. S., Law, P. Y. T., Leung, G. M., Peiris, M., Poon, L. L. M., Cheng, S. M. S., Chang, L. D. J., Krishnan, P., Ng, D. Y. M., Liu, G. Y. Z., Hui, M. M. Y., Ho, S. Y., Wen, S., Sia, S. F., Choy, K.-T., … Yung, L. (2022). Transmission of SARS-CoV-2 delta variant (AY.127) from pet hamsters to humans, leading to onward human-to-human transmission: a case study. *The Lancet*, *399*, 1070–1078.

Katoh, K., & Standley, D. M. (2013). MAFFT multiple sequence alignment software version 7: improvements in performance and usability. *Molecular Biology and Evolution*, *30*, 772–780.

Kim, Y. I., Kim, S. G., Kim, S. M., Kim, E. H., Park, S. J., Yu, K. M., Chang, J. H., Kim, E. J., Lee, S., Casel, M. A. B., Um, J., Song, M. S., Jeong, H. W., Lai, V. D., Kim, Y., Chin, B. S., Park, J. S., Chung, K. H., Foo, S. S., … Choi, Y. K. (2020). Infection and rapid transmission of SARS-CoV-2 in ferrets. *Cell Host & Microbe*, *27*(5), 704–709.e2.

Larsen, C. S., & Paludan, S. R. (2020). Corona's new coat: SARS-CoV-2 in Danish minks and implications for travel medicine. *Travel Medicine and Infectious Disease*, *38*, 101922.

Lazarevic, I., Pravica, V., Miljanovic, D., & Cupic, M. (2021). Immune evasion of SARS-CoV-2 emerging variants: What have we learnt so far? *Viruses*, *13*(7), 1192.

Letunic, I., & Bork, P. (2019). Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Research*, *47*, W256–W259.

Leung, K., Shum, M. H., Leung, G. M., Lam, T. T., & Wu, J. T. (2021). Early transmissibility assessment of the N501Y mutant strains of SARS-CoV-2 in the United Kingdom. *Euro Surveillance*, *26*, 2002106.

Liu, P., Jiang, J. Z., Wan, X. F., Hua, Y., Li, L., Zhou, J., Wang, X., Hou, F., Chen, J., Zou, J., & Chen, J. (2020). Are pangolins the intermediate host of the 2019 novel coronavirus (SARS-CoV-2)? *PLoS Pathogens*, *16*, e1008421.

Luan, J., Jin, X., Lu, Y., & Zhang, L. (2020). SARS-CoV-2 spike protein favors ACE2 from Bovidae and Cricetidae. *Journal of Medical Virology*, *92*, 1649–1656.

Lytras, S., Xia, W., Hughes, J., Jiang, X., & Robertson, D. L. (2021). The animal origin of SARS-CoV-2. *Science*, *373*, 968–970.

Mallapaty, S. (2020). COVID mink analysis shows mutations are not dangerous - yet. *Nature*, *587*, 340–341.

McAloose, D., Laverack, M., Wang, L., Killian, M. L., Caserta, L. C., Yuan, F., Mitchell, P. K., Queen, K., Mauldin, M. R., Cronk, B. D., Bartlett, S. L., Sykes, J. M., Zec, S., Stokol, T., Ingerman, K., Delaney, M. A., Fredrickson, R., Ivancic, M., Jenkins-Moore, M., … Diel, D. G. (2020). From people to panthera: Natural SARS-CoV-2 infection in tigers and lions at the bronx zoo. *mBio*, *11*(5), e02220.

Mlcochova, P., Kemp, S., Dhar, M. S., Papa, G., Meng, B., Ferreira, I., Datir, R., Collier, D. A., Albecka, A., Singh, S., Pandey, R., Brown, J., Zhou, J., Goonawardane, N., Mishra, S., Whittaker, C., Mellan, T., Marwal, R., Datta, M., … Gupta, R. K. (2021). SARS-CoV-2 B.1.617.2 Delta variant replication and immune evasion. *Nature*, *599*, 114–119.

Nguyen, L. T., Schmidt, H. A., von Haeseler, A., & Minh, B. Q. (2015). IQ-TREE: a fast and effective stochastic algorithm for estimating maximum-likelihood phylogenies. *Molecular Biology and Evolution*, *32*, 268–274.

Oude Munnink, B. B., Sikkema, R. S., Nieuwenhuijse, D. F., Molenaar, R. J., Munger, E., Molenkamp, R., van der Spek, A., Tolsma, P., Rietveld, A., Brouwer, M., Bouwmeester-Vincken, N., Harders, F., Hakze-van der Honing, R., Wegdam-Blans, M. C. A., Bouwstra, R. J., GeurtsvanKessel, C., van der Eijk, A. A., Velkers, F. C., Smit, L. A. M., … Koopmans, M. P. G. (2021). Transmission of SARS-CoV-2 on mink farms between humans and mink and back to humans. *Science*, *371*, 172–177.

Patro, L. P. P., Sathyaseelan, C., Uttamrao, P. P., & Rathinavelan, T. (2021a). The evolving proteome of SARS-CoV-2 predominantly uses mutation combination strategy for survival. *Computational and Structural Biotechnology Journal*, *19*, 3864–3875.

Patro, L. P. P., Sathyaseelan, C., Uttamrao, P. P., & Rathinavelan, T. (2021b). Global variation in SARS-CoV-2 proteome and its implication in pre-lockdown emergence and dissemination of 5 dominant SARS-CoV-2 clades. *Infection, Genetics and Evolution*, *93*, 104973.

Plante, J. A., Liu, Y., Liu, J., Xia, H., Johnson, B. A., Lokugamage, K. G., Zhang, X., Muruato, A. E., Zou, J., Fontes-Garfias, C. R., Mirchandani, D., Scharton, D., Bilello, J. P., Ku, Z., An, Z., Kalveram, B., Freiberg, A. N., Menachery, V. D., Xie, X., … Shi, P. Y. (2021). Spike mutation D614G alters SARS-CoV-2 fitness. *Nature*, *592*, 116–121.

Sailleau, C., Dumarest, M., Vanhomwegen, J., Delaplace, M., Caro, V., Kwasiborski, A., Hourdel, V., Chevaillier, P., Barbarino, A., Comtet, L., Pourquier, P., Klonjkowski, B., Manuguerra, J. C., Zientara, S., & Le Poder, S. (2020). First detection and genome sequencing of SARS-CoV-2 in an infected cat in France. *Transboundary and Emerging Diseases*, *67*, 2324–2328.

Samavati, L., & Uhal, B. D. (2020). ACE2, much more than just a receptor for SARS-COV-2. *Frontiers in Cellular and Infection Microbiology*, *10*, 317.

Sharun, K., Dhama, K., Pawde, A. M., Gortazar, C., Tiwari, R., Bonilla-Aldana, D. K., Rodriguez-Morales, A. J., de la Fuente, J., Michalak, I., & Attia, Y. A. (2021). SARS-CoV-2 in animals: potential for unknown reservoir hosts and public health implications. *The Veterinary Quarterly*, *41*, 181–201.

Shi, J., Wen, Z., Zhong, G., Yang, H., Wang, C., Huang, B., Liu, R., He, X., Shuai, L., Sun, Z., Zhao, Y., Liu, P., Liang, L., Cui, P., Wang, J., Zhang, X., Guan, Y., Tan, W., Wu, G., … Bu, Z. (2020). Susceptibility of ferrets, cats, dogs, and other domesticated animals to SARS-coronavirus 2. *Science*, *368*, 1016–1020.

Sit, T. H. C., Brackman, C. J., Ip, S. M., Tam, K. W. S., Law, P. Y. T., To, E. M. W., Yu, V. Y. T., Sims, L. D., Tsang, D. N. C., Chu, D. K. W., Perera,

R., Poon, L. L. M., & Peiris, M. (2020). Infection of dogs with SARS-CoV-2. *Nature*, *586*, 776–778.

Uttamrao, P. P., Sathyaseelan, C., Patro, L. P. P., & Rathinavelan, T. (2021). Revelation of potent epitopes present in unannotated ORF antigens of SARS-CoV-2 for epitope-based polyvalent vaccine design using immunoinformatics approach. *Frontiers in Immunology*, *12*, 692937.

Volz, E., Hill, V., McCrone, J. T., Price, A., Jorgensen, D., O'Toole, A., Southgate, J., Johnson, R., Jackson, B., Nascimento, F. F., Rey, S. M., Nicholls, S. M., Colquhoun, R. M., da Silva Filipe, A., Shepherd, J., Pascall, D. J., Shah, R., Jesudason, N., Li, K., … Connor, T. R. (2021). Evaluating the effects of SARS-CoV-2 spike mutation D614G on transmissibility and pathogenicity. *Cell*, *184*(1), 64–75.e11.

Yang, T. J., Yu, P. Y., Chang, Y. C., & Hsu, S. D. (2021). D614G mutation in the SARS-CoV-2 spike protein enhances viral fitness by desensitizing it to temperature-dependent denaturation. *The Journal of Biological Chemistry*, *297*, 101238.

Zhang, L., Jackson, C. B., Mou, H., Ojha, A., Peng, H., Quinlan, B. D., Rangarajan, E. S., Pan, A., Vanderheiden, A., Suthar, M. S., Li, W., Izard, T., Rader, C., Farzan, M., & Choe, H. (2020). SARS-CoV-2 spike-protein D614G mutation increases virion spike density and infectivity. *Nature Communications*, *11*, 6013.

SUPPORTING INFORMATION

Additional supporting information may be found in the online version of the article at the publisher's website.

How to cite this article: Magateshvaren Saras, M. A., Patro, L. P., Uttamrao, P. P., & Rathinavelan, T. (2022). Geographical distribution of SARS-CoV-2 amino acids mutations and the concomitant evolution of seven distinct clades in non-human hosts. *Zoonoses and Public Health*, *00*, 1–10. https://doi.org/10.1111/zph.12971