SARS-CoV-2 NSP14 governs mutational instability and assists in making new SARS-CoV-2 variants

Sk. Sarif Hassan[a,*], Tanishta Bhattacharya[**,b], Debaleena Nawn[**,c], Ishana Jha[d], Pallab Basu[e,f], Elrashdy M. Redwan[g], Kenneth Lundstrom[h], Debmalya Barh[h,j], Bruno Silva Andrade[k], Murtaza M. Tambuwala[l,*], Alaa A. Aljabali[m], Altijana Hromić-Jahjefendić[n], Wagner Baetas-da-Cruz[o], Vladimir N. Uversky[p,*]

[a]*Department of Mathematics, Pingla Thana Mahavidyalaya, Maligram, Paschim Medinipur, 721140, West Bengal, India*
[b]*Department of Biological Sciences, Indian Institute of Science Education and Research, Berhampur, IISER Berhampur Transit campus (Govt. ITI Building), Engg. School Junction, Berhampur, 760010, Odisha, India*
[c]*Agricultural and Ecological Research Unit, Indian Statistical Institute, Kolkata, 203 B T Road, Kolkata, 700108, West Bengal, India*
[d]*Department of Bioinformatics, Pondicherry University, Chinna Kalapet, Kalapet, Puducherry 605014, India*
[e]*School of Physics, University of the Witwatersrand, Johannesburg, Braamfontein 2000, 721140, South Africa*
[f]*Woxsen School of Sciences, Woxsen University, Telangana, 500 033, India*
[g]*Biological Science Department, Faculty of Science, King Abdulaziz University, Jeddah, Saudi Arabia, Therapeutic and Protective Proteins Laboratory, Protein Research Department, Genetic Engineering and Biotechnology Research Institute, City of Scientific Research and Technological Applications, New Borg EL-Arab, 21934, Alexandria, Egypt*
[h]*PanTherapeutics, Rte de Lavaux 49, CH1095 Lutry, Switzerland*
[i]*Institute of Integrative Omics and Applied Biotechnology (IIOAB), Nonakuri, Purba Medinipur, 721172, India*
[j]*Department of Genetics, Ecology and Evolution, Institute of Biological Sciences, Federal University of Minas Gerais, Belo Horizonte, 31270-901, Brazil*
[k]*Laboratory of Bioinformatics and Computational Chemistry, Department of Biological Sciences, State University of Southwest of Bahia (UESB), Jequié 45083-900, Brazil*
[l]*Lincoln Medical School, University of Lincoln, Brayford Pool Campus, Lincoln LN6 7TS, UK*
[m]*Department of Pharmaceutics and Pharmaceutical Technology, Faculty of Pharmacy, Yarmouk University, Irbid 21163, Jordan*
[n]*Department of Genetics and Bioengineering, Faculty of Engineering and Natural Sciences, International University of Sarajevo, Hrasnicka cesta 15, 71000 Sarajevo, Bosnia and Herzegovina*
[o]*Centre for Experimental Surgery, Translational Laboratory in Molecular Physiology, College of Medicine, Federal University of Rio de Janeiro (UFRJ), Rio de Janeiro, Brazil*
[p]*Department of Molecular Medicine, Morsani College of Medicine, University of South Florida, Tampa, FL 33612, USA*

Abstract

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), the rapidly evolving RNA virus behind the COVID-19 pandemic, has spawned numerous variants since its 2019 emergence. The multifunctional NSP14 enzyme, possessing exonuclease and mRNA capping capabilities, serves as a key player. Notably, single and co-occurring mutations within NSP14 significantly influence replication fidelity and drive variant diversification. This study comprehensively examines 120 co-mutations, 68 unique mutations, and 160 conserved residues across NSP14 homologs, shedding light on their implications for phylogenetic patterns, pathogenicity, and residue interactions. Quantitative physicochemical analysis categorizes 3953 NSP14 variants into three clusters, revealing genetic diversity. This research underscores the dynamic nature of SARS-CoV-2 evolution, primarily governed by NSP14 mutations. Understanding these genetic dynamics provides valuable insights for therapeutic and vaccine development.

Keywords: SARS-CoV-2, NSP14, Co-mutations-flow, Invariant Residues, Clustering, Phylogeny, RdRp

1. Introduction

The emergence of SARS-CoV-2 in December 2019 had a profound global impact. This virus, a member of the Coronaviridae family, is a betacoronavirus characterized by its positive-sense, single-stranded RNA genome [1, 2]. Sequence analysis revealed striking similarities between SARS-CoV-2 and its predecessors, MERS-CoV (2013) and SARS-CoV (2001) [3]. Upon infection, SARS-CoV-2 can manifest with a wide spectrum of symptoms, ranging from pneumonia to severe acute respiratory syndrome [4]. Its genome encodes four structural proteins—Spike (S), Envelope (E), Membrane (M), and Nucleocapsid (N)-and sixteen non-structural proteins (NSP1 to NSP16), primarily involved in replication and protease activity [5, 6]. Due to its high transmissibility and devastating global impact, extensive research on the SARS-CoV-2 genome has led to the creation of valuable databases such as GISAID and NextStrain, providing crucial insights into its genetic makeup [7, 8, 9].

*Corresponding author
**Authors contributed equally
Email addresses: sksarifhassan@pinglacollege.ac.in (Sk. Sarif Hassan), tanishta18@iiserbpr.ac.in (Tanishta Bhattacharya[**,]), debaleena.nawn@gmail.com (Debaleena Nawn[**,]), ishanajroy@gmail.com (Ishana Jha), pallabbasu@gmail.com (Pallab Basu[e,]), lradwan@kau.edu.sa (Elrashdy M. Redwan), lundstromkenneth@gmail.com (Kenneth Lundstrom), dr.barh@gmail.com (Debmalya Barh[h,]), bandrade@uesb.edu.br (Bruno Silva Andrade), mtambuwala@lincoln.ac.uk (Murtaza M. Tambuwala), alaaj@yu.edu.jo (Alaa A. Aljabali), ahromic@ius.edu.ba (Altijana Hromić-Jahjefendić), wagner.baetas@gmail.com (Wagner Baetas-da-Cruz), vuversky@usf.edu (Vladimir N. Uversky)

The SARS-CoV-2 genome, roughly 30 Kb in length, is dominated by two polyproteins, pp1a and pp1ab, which are subsequently cleaved into non-structural proteins. Among these, RNA-dependent RNA polymerase (RdRp) is central to RNA synthesis, forming a core complex with nsp7, nsp8, and nsp12, while nsp14 interacts with this complex, wielding essential proofreading activity [6, 10]. NSP14, consisting of approximately 527 amino acids, possesses a bifunctional role. Its N-terminal region harbors a 3'-5' exonuclease responsible for proofreading during replication, ensuring genomic fidelity and reducing error rates [11, 12, 13]. The C-terminal N7-methyltransferase domain plays a critical role in mRNA capping and evading host immune responses [14]. Mutations in the DxG domain within this region can abolish RNA capping activity without affecting exonuclease function [15]. While RdRp facilitates viral replication, post-transcriptional processes involve capping and polyadenylation to protect the genome from degradation, evade antivirals, and enable efficient host translation. This is orchestrated by non-structural proteins NSP13, NSP14, and NSP16, with NSP14 typically associating with its accessory protein, NSP10 [16, 17].

The NSP10/14 complex plays a pivotal role in maintaining replication fidelity, controlling mutation rates, and influencing resistance to antiviral drugs [18]. Recent research highlights compounds like SGC0946 and SGC8158 that selectively target SARS-CoV-2 NSP14 by blocking SAM and RNA-binding sites [19].

RNA viruses, known for their high mutation rates, create diverse viral strains. These mutations can either enhance virulence, enable drug resistance, or go unnoticed without impacting transmission. The presence of a proofreading mechanism, like that found in SARS-CoV-2, can reduce mutation rates compared to other viruses, estimated at approximately $10^{-}6$ mutations per site/cycle [1]. Mutations in NSP14, particularly in the ExoN domain, contribute to increased genome-wide mutational loads [9]. Specific mutations, such as the shift from cytosine to uracil, can affect host inflammatory cytokine production [20]. Mutations in the active sites and zinc finger motifs of the Exonuclease domain can lead to lethal phenotypes [21, 11]. The significance of understanding these mutations lies in their impact on diagnostic assays, vaccine development, and viral transmissibility [22, 23]. Targeting NSP14, which maintains genomic stability, offers potential therapeutic strategies against SARS-CoV-2 [24]. Beyond its role in proofreading, NSP14 acts as a translation inhibitor in host cells, a characteristic conserved across human coronaviruses [23]. Recent studies have identified natural compounds that can target several key residues within NSP14 as potential targets for therapeutic interventions [25]. Conservation across coronaviruses is evident, with the NSP14 of SARS-CoV-2 sharing high amino acid similarity with its counterparts in SARS-CoV and MERS-CoV [23]. The ExoN and N7-MTase domains play vital roles in inhibiting host antiviral responses, and mutations in these domains can trigger host antiviral responses [23].

This study explores unique mutations in NSP14 across 13 different geographic locations and assesses their impact, differentiating between neutral and deleterious mutations using the PredSNP server [26].

2. Data Specifications and Methods

2.1. Data

A total of 1,182,629 SARS-CoV-2 NSP14 sequences, devoid of any ambiguous characters, were sourced from the GISAID-Virusurf database. Remarkably, among this extensive dataset, only 3,953 sequences, equivalent to a mere 0.33%, exhibited unique and distinct variations, representing the global diversity of SARS-CoV-2 NSP14. For reference, the NSP14 sequence denoted as YP_009725309 was retrieved from the NCBI.

2.2. Methods

2.2.1. Analyze sequence variation

Single mutations in all the 3953 unique NSP14 sequences were determined using the Virus Pathogen Resource ViPR by inputting Fasta file of NSP14 sequences [27].

Furthermore, the predicted effect on pathogenicity of all the mutations was analyzed with PredictSNP and PhD-SNP [26, 28]. Note that PredictSNP web server makes a consensus based on other prediction tools such as MAPP, PolyPhen-1 and PolyPhen-2, SIFT, SNAP, and PANTHER. Therefore, a degree of accuracy is expected to be ensured. In addition, co-occurrence of mutations in NSP14 variants were also detected by the CoVal database.

2.2.2. Prediction of RNA-interacting residues

Prediction of RNA-interacting residues in the SARS-CoV-2 NSP14 protein was made by the webserver Pprint. The webserver takes the amino acid sequence and automatically generates the evolutionary profile of whole sequence by running PSI-BLAST, generates support vector machine (SVM) pattern from this position-specific scoring matrix (PSSM) profile and then, predicts RNA interacting residues using SVM model[29]. The probability of correct prediction directly depends on the threshold, which by default set as -0.2. In this study, default threshold was used [29]. The red coloured residues are predicted as RNA-interacting and blue coloured residues are predicted as non RNA-interacting.

2.2.3. Frequency distribution of amino acids and clustering

The frequency distribution of each amino acid present in a NSP14 sequence was determined using standard bioinformatics routine in *henson2004matlab*. For each NSP14 sequence, a twenty-dimensional frequency-vector considering the frequency of standard twenty amino acids can be obtained. The distance (Euclidean metric) between any two pairs of frequency vectors was calculated for each pair of NSP14 variants.

2.2.4. Density-Based Clustering

Density-Based Spatial Clustering of Applications with Noise (DBSCAN) is one the most popular density-based clustering algorithm [30]. Twenty dimensional frequency vectors for all 3953 NSP14 sequences were considered as the dataset for clustering of 3953 SARS-CoV-2 NSP14 sequences. DBSCAN is a density-based clustering non-parametric algorithm: given a set of points in some space, it groups together points that are closely packed together. For the plotting purpose, we project the resulting clusters to a plane spanned by two principal components [30].

3. Results

3.1. Single Point Mutations in NSP14 Variants

Among the 527 amino acid residue positions, a total of 962 single mutations were identified within 3,953 unique NSP14 variants, as detailed in Tables (11 - 13). Notably, among these mutations, 398 were located within the ExoN domain, and 348 mutations were situated in the N7-MTase domain of the NSP14 sequence. Intriguingly, multiple mutations with frequencies ranging from 1 to 7 were observed at numerous residue positions, with seven distinct mutations documented at residues 132, 204, 211, and 496. A visual representation of mutation frequencies at each location is depicted in Figure 1.

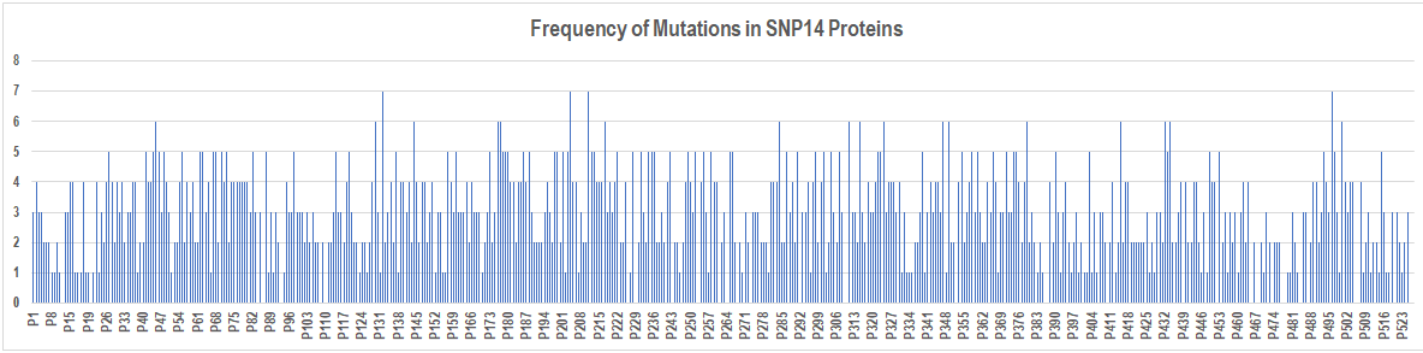

Figure 1: Total frequency of mutations per mutant residues of NSP14 reference sequence (YP_009725309)

Furthermore, a comprehensive analysis revealed a total of 414 deleterious mutations and 548 neutral mutations, classified based on pathogenicity, as outlined in Tables (11 - 13).

Notably, a variety of unique mutations were discovered across different geographic regions, as presented in Table 1. It's worth mentioning that the highest number of unique single mutations within NSP14 variants were observed in the United States and the United Kingdom (Table 2). Additionally, it was observed that nearly half of the identified unique mutations were deleterious in the USA, UK, and India (Table 2).

Table 1: Unique single point mutations in NSP14 Variants in UK, USA, India, Canada, Switzerland, France, Spain, Mexico, Italy, Japan, and Denmark

UK			USA			India	Canada	Switzerland	France	Spain
H427L	H148Q	L117P	I338L	F436I	A281G	A504Y	L177V	W227L	M241L	L411M
I294M	I242L	N306K	L406I	N129G	D331Y	D464G	Q145P	D30E	S407A	F377Y
D496H	I294S	Q313E	E204Q	N129Y	G59V	F367C		Y260F	T277A	
G417S	K433N	S28N	K175R	D291G	K311T	H455S	**Italy**	**Mexico**	**Japan**	**Denmark**
R391S	R213S	V4A	V182A	D515A	N67K	S418V	W348R	S230T	A307P	Q513R
D324H	R525I		D179A	F89C	N71T	Y51C				
E365G	Y361F		F350S	I332L	P451R					
I166S	C382S		L322V	M57K	S369Y					
C216L	I299M		N129A	S194C						

Table 2: Total number and percentage of deleterious and neutral mutations in different geo-locations

Unique mutations	UK	USA	India	Canada	Italy	Switzerland	Mexico	France	Japan	Spain	Denmark
Total mutations	23	26	6	4	1	1	1	3	1	1	1
Total Neutral mutations	13	13	3	2	0	1	1	1	1	0	1
Total Deleterious mutations	10	13	3	2	1	0	0	2	0	1	0
Neutral (%)	56.52	50	50	50	0	100	100	33.3	100	0	100
Deleterious (%)	43.48	50	50	50	100	0	0	66.6	0	100	0

3.1.1. Invariant Residues across NSP14 of various coronaviruses

Table 3 presents the invariant residues within the NSP14 reference sequence, emphasizing their consistency across 16 different coronaviruses originating from various hosts. The sequence homology analysis was conducted using Clustal-Omega, encompassing the following coronaviruses: Rousettus bat coronavirus HKU9 (YP_009924395.1), Betacoronavirus England (YP_009944304.1), Pipistrellus bat coronavirus HKU5 (YP_009944351.1), Tylonycteris bat coronavirus HKU4 (YP_009944322.1), SARS coronavirus Tor2 (NP_828871.1), Rabbit coronavirus HKU14 (YP_009924421.1), Murine hepatitis virus (YP_009915687.1), Rat coronavirus Parker (YP_009924380.1), Murine hepatitis virus (YP_009924354.1), Duck coronavirus (YP_009825026.1), Canada goose coronavirus (YP_009755922.1), Human coronavirus HKU1 (YP_460021.1), Murine hepatitis virus strain JHM (YP_209241.1), Bat coronavirus CDPHE15/USA/2006 (YP_008439224.1), Turkey coronavirus (YP_001941187.1), and the NSP14 reference sequence (YP_009725309.1).

These invariant residues signify regions of the NSP14 protein that exhibit a remarkable degree of conservation across diverse coronaviruses, regardless of their host origins.

Table 3: Invariant residues across NSP14 of various coronaviruses

Invariant Residue	**Invariant Residue**	**Invariant Residue**	**Invariant Residue**	**Invariant Residue**	**Invariant Residue**
7 (L)	112 (S)	226 (C)	280 (L)	400 (R)	452 (C)
9 (K)	114 (G)	229 (H)	281 (A)	402 (D)	466 (V)
11 (C)	118 (V)	234 (D)	286 (F)	403 (T)	468 (L)
20 (P)	123 (G)	237 (Y)	292 (W)	404 (R)	473 (C)
34 (K)	141 (P)	238 (N)	296 (Y)	410 (N)	474 (I)
51 (Y)	142 (P)	239 (P)	297 (P)	411 (L)	475 (T)
56 (S)	143 (G)	243 (D)	299 (I)	413 (G)	477 (C)
59 (G)	146 (F)	245 (Q)	302 (E)	414 (C)	478 (N)
60 (F)	148 (H)	246 (Q)	306 (N)	416 (G)	480 (G)
73 (F)	149 (L)	247 (W)	310 (R)	417 (G)	481 (G)
75 (T)	152 (L)	248 (G)	331 (D)	418 (S)	482 (A)
79 (A)	159 (W)	251 (G)	332 (I)	419 (L)	483 (V)
83 (V)	163 (R)	253 (L)	333 (G)	420 (Y)	484 (C)
84 (R)	166 (I)	256 (N)	334 (N)	421 (V)	487 (A)
86 (W)	169 (M)	257 (H)	335 (P)	422 (N)	488 (A)
88 (G)	172 (D)	261 (C)	336 (K)	424 (H)	491 (Y)
89 (F)	186 (W)	264 (H)	352 (D)	425 (A)	498 (Y)
90 (D)	191 (E)	268 (H)	355 (P)	426 (F)	499 (N)
92 (E)	192 (L)	269 (V)	368 (Y)	427 (H)	505 (G)
95 (H)	193 (T)	270 (A)	379 (D)	428 (T)	506 (F)
102 (G)	197 (Y)	271 (S)	380 (G)	436 (F)	509 (W)
103 (T)	198 (F)	273 (D)	388 (N)	439 (L)	514 (F)
104 (N)	200 (K)	274 (A)	389 (V)	440 (K)	518 (N)
106 (P)	202 (G)	276 (M)	392 (Y)	443 (P)	519 (L)
108 (Q)	207 (C)	277 (T)	393 (P)	444 (F)	520 (W)
110 (G)	210 (C)	278 (R)	398 (V)	445 (F)	
111 (F)	214 (A)	279 (C)	399 (C)	447 (Y)	

A total of 160 invariant residue positions, identified through amino acid sequence homology comparisons with NSP14 sequences from various coronaviruses, were found in relation to the SARS-CoV-2 NSP14 reference sequence. Interestingly, it was observed that despite their invariance across other coronaviruses, several mutations were detected at most of these invariant residue positions within the different SARS-CoV-2 NSP14 sequences, as detailed in Table 4.

Table 4: Type of mutations detected in SARS-CoV-2 NSP14 variants at the invariant residues (Table-3)

Invariant Residue	Type of mutation	Invariant Residue	Type of mutation	Invariant Residue	Type of mutation
7 (L)	Deleterious	214 (A)	Deleterious	404 (R)	Deleterious
34 (K)	Deleterious, Neutral	234 (D)	Deleterious	411 (L)	Deleterious
51 (Y)	Neutral	238 (N)	Deleterious	413 (G)	Deleterious
59 (G)	Deleterious	239 (P)	Deleterious	414 (C)	Deleterious
75(T)	Deleterious	245 (Q)	Deleterious	416 (G)	Deleterious
79 (A)	Deleterious	247 (W)	Deleterious	417 (G)	Neutral
83 (V)	Deleterious	248 (G)	Deleterious	418 (S)	Deleterious
88 (G)	Deleterious	251 (G)	Deleterious	420 (Y)	Deleterious
89 (F)	Deleterious	269 (V)	Deleterious	421 (V)	Deleterious, Neutral
95 (H)	Deleterious	274 (A)	Deleterious	424 (H)	Deleterious
103 (T)	Deleterious	276 (M)	Deleterious	425 (A)	Deleterious
111 (F)	Deleterious	277 (T)	Deleterious	427 (H)	Deleterious, Neutral
112 (S)	Deleterious	278 (R)	Neutral	428 (T)	Deleterious
114 (G)	Deleterious	281 (A)	Deleterious	436 (F)	Deleterious
118 (V)	Deleterious, Neutral	296 (Y)	Deleterious	440 (K)	Neutral
141 (P)	Deleterious	297 (P)	Deleterious	443 (P)	Deleterious
142 (P)	Deleterious	299 (I)	Deleterious	444 (F)	Deleterious
143 (G)	Deleterious	306 (N)	Deleterious	447 (Y)	Deleterious
148 (H)	Deleterious	331 (D)	Deleterious	466 (V)	Neutral
149 (L)	Deleterious	332 (I)	Deleterious, Neutral	474 (I)	Neutral
152 (L)	Deleterious, Neutral	336 (K)	Deleterious	475 (T)	Deleterious
159 (W)	Deleterious	355 (P)	Deleterious	481 (G)	Deleterious, Neutral
163 (R)	Deleterious	379 (D)	Deleterious, Neutral	482 (A)	Deleterious
166 (I)	Deleterious	380 (G)	Deleterious	488 (A)	Deleterious
169 (M)	Deleterious	388 (N)	Deleterious	491 (Y)	Deleterious
191 (E)	Deleterious	389 (V)	Deleterious	498 (Y)	Deleterious
193 (T)	Neutral	393 (P)	Deleterious	506 (F)	Deleterious
197 (Y)	Deleterious	398 (V)	Deleterious	509 (W)	Deleterious
198 (F)	Deleterious	399 (C)	Deleterious	514 (F)	Deleterious
200 (K)	Deleterious	403 (T)	Deleterious		

Table 4 indeed reveals a noteworthy observation: nearly all of the invariant residues identified in Table 3, which were considered invariant across various coronaviruses, exhibited deleterious mutations within the SARS-CoV-2 NSP14 variants. This finding underscores the potential significance of these invariant positions in the functional and structural integrity of NSP14, as mutations in these positions are more likely to have detrimental effects on the virus.

3.1.2. Invariant Residues across NSP14 Variants

In Table 5, we presented a list of 110 "Invariant residues" of the SARS-CoV-2 NSP14. These are amino acid residues that did not exhibit any single mutations in the unique NSP14 variants considered in this study, highlighting their remarkable stability and lack of variation within this specific dataset.

Table 5: 110 invariant residues positions in SARS-CoV-2 NSP14 variants (amino acid residues marked in bold font are invariant across all 16 NSP14 sequences of coronaviruses as found in Table 3)

AA position of Invariant Residues in NSP14 Variants								
AA Position	***AA Position***	***AA Position***	***AA Position***	***AA Position***	***AA Position***	***AA Position***	***AA Position***	***AA Position***
8	**73**	**110**	**226**	**273**	340	**419**	470	**518**
9	**84**	**123**	**229**	**279**	351	**422**	**473**	**519**
11	**86**	126	**237**	**280**	**352**	**426**	**477**	**520**
12	**90**	127	**243**	285	362	**439**	**478**	523
18	**92**	**146**	**246**	**286**	**368**	**445**	479	526
19	93	170	**253**	**292**	385	446	**480**	527
20	94	**172**	**256**	**302**	386	**452**	**483**	
25	**102**	**186**	**257**	309	387	454	**484**	
38	**104**	**192**	**261**	**310**	**392**	456	**487**	
56	105	**202**	**264**	**333**	**400**	458	492	
60	**106**	**207**	**268**	**334**	401	465	**499**	
65	**108**	209	**270**	**335**	**402**	467	**505**	
66	109	**210**	**271**	337	**410**	**468**	517	

Based on the information presented in Table 5, it was deduced that 71 amino acid residues, marked in bold font, remained invariant across all 16 NSP14 sequences of various coronaviruses. In contrast, 37 amino acid residues (e.g., 8,

12, 18, 19, and others), which were not bolded in Table 5, were observed to be invariant in the unique 3953 SARS-CoV-2 NSP14 variants but exhibited variability across the other 16 different NSP14 coronavirus sequences, as detailed in Table 3.

Furthermore, the analysis of Table 5 led to the identification of double, triple, and quadruple invariant regions, as outlined in Table 6. It was noted that only two quadruple invariant regions, CNLG (amino acid positions: 477-480) and YNLW (amino acid positions: 517-520), contained both the PPNN ordered residues. This finding highlights the uniqueness and potential functional significance of these specific regions within the NSP14 protein.

Table 6: Invariant regions of double, triple, and quadruple sizes and their polarity (P and N denote polar and non-polar residues, respectively). Yellow and blue marked regions belong to the Exon and N7-MTase domains of SARS-CoV-2 NSP14.

Double Residues	Polar-Non-Polarity	Double Residues	Polar-Non-Polarity	Triple Residues	Polar-Non-Polarity	Quadruple Residues	Polar-Non-Polarity
FK(8-9)	NP	CL(279-280)	PN	LHP(18-20)	NPN	CNLG(477-480)	PPNN
CS(11-12)	PP	CF(285-286)	PN	EGC(92-94)	PNP	YNLW(517-520)	PPNN
QV(65-66)	PN	CR(309-310)	PP	NLP(104-106)	PNN		
DT(126-127)	PP	YD(351-352)	PP	QLG(108-110)	PNN		
LC(209-210)	NP	FY(445-446)	NP	GNP(333-335)	NPN		
NH(256-257)	PP	PL(467-468)	NN	WNC(385-387)	NPP		
AS(270-271)	NP	VC(483-484)	NP	RFD(400-402)	PNP		
		LQ(526-527)	NP				

3.2. Co-occurring mutations in NSP14 variants

Beyond the single non-synonymous mutations, it was observed that a limited number of co-occurring mutations were present in NSP14 variants originating from various geographic locations. These co-occurring mutations suggest potential interactions or synergistic effects among specific amino acid residues within the NSP14 protein, warranting further investigation into their functional implications.

3.2.1. Co-occurring mutations and their classification based on pathogenicity

Table 7 effectively tabulated the co-occurring mutations found in NSP14 variants from various geographic locations. These co-occurring mutations were distinguished as "unique co-occurring mutations" because they were exclusively identified with multiple frequencies (≥ 2) in their respective geographic locations. Notably, it was observed that all co-occurring mutations listed in Table 7 were also previously detected as single mutations, as indicated in Tables 10-12. Conversely, there were single mutations that had not yet been identified as co-occurring mutations, suggesting distinct mutation patterns and dynamics within the NSP14 variants.

Table 7: List of unique co-occurring mutations in different geo-locations. All the deleterious mutations occurring (such as G6R;L177F, H373Y;L157F and so on) together were marked bold.

Country	*Mutations*	*Frequency*	*Date first collected*	*Country*	*Mutations*	*Frequency*	*Date first collected*	*Country*	*Mutations*	*Frequency*	*Date first collected*
UK	Q22H;C208S	248	21-06-2020	India	**F217V;Y511H**	2	25-06-2020	USA	T16I;N129D	26	27-11-2020
	M58I;Y260C	33	08-05-2020		N176D;H455Y	2	06-02-2021		N129D;S255I	25	06-03-2021
	E347G;E453D	28	21-03-2021	Canada	V381L;H455Y	9	01-03-2021		K147R;P412H	23	19-02-2021
	V182L;A394V	28	22-06-2021		N129D;S374F	2	27-11-2020		N129D;E204D	21	15-11-2020
	V120A;A394V	19	31-05-2021		P443S;P451S	2	24-03-2021		K304T;P451S	16	14-03-2021
	P24L;H26Y	12	03-04-2020	Denmark	S374A;L493F	5	27-07-2020		P142S;A353V	13	26-11-2020
	V125F;M501I	10	24-08-2020		A119V;V236F	4	16-11-2020		N129D;T219I	12	21-03-2021
	S369F;A394V	7	07-07-2021		P46S;G248V	3	26-04-2021		V381L;D496Y	12	05-03-2021
	S221P;A471V	5	02-11-2020		P140L;A371V	2	21-12-2020		**M169I;P412H**	11	17-03-2021
	C208S;A394V	4	14-07-2021		V263F;P451S	2	11-01-2021		A220V;M501I	10	20-03-2021
	Q22H;C208S;S418I	4	10-11-2020	France	**A119S;M169I**	8	2021-05		M153V;K155R	10	07-04-2021
	R163C;S461L	4	18-04-2020		**L157F;H373Y**	3	15-03-2021		N129D;H486R	8	30-01-2021
	R212K;A323S	4	05-10-2020	Italy	M49V;T524I	2	20-04-2021		N129D;L366F	8	28-02-2021
	A371V;L493F	3	03-05-2021	Japan	P43L;G251S	83	2020-11		N129D;P158S	8	24-03-2021
	K349N;M501I	3	29-10-2020		P43L;V510A	45	2020-07		V125F;I166V	7	03-07-2020
	R213L;V287F	3	30-09-2020		P43L;Q343L	21	2020-09		N129D;R212K	6	12-10-2020
	H26Y;R525K	2	09-12-2020		P43L;A138V	15	2020-08		V14L;V381L	6	18-11-2020
	L157F;A225V	2	24-03-2020		K311N;P412H	14	2021-01		N129D;A360S	5	03-09-2020
	M62L;N63K	2	14-08-2020		P43L;P140S	12	2020-11		N129D;K304R	5	29-12-2020
	N238S;V421I	2	09-09-2020		P43L;L177F	7	06-12-2020		N129D;T524I	5	2021-03
	P140L;A344S	2	22-11-2020		P43L;I74T	6	2020-12		T16I;M58I	5	03-09-2020
	P142S;D222Y	2	03-01-2021		P43L;T250I	6	2020-11		T31I;V341A	5	18-03-2021
	Q22H;C208S;T524I	2	12-11-2020		P43L;V328F	6	2020-12		D48N;N129D	4	11-01-2021
	Q22H;E36G	2	27-09-2020		P43L;H486L	3	2021-01		M49V;N129D	4	05-11-2020
	Q22H;V381I	2	09-10-2020		P43L;M501I	3	2020-10		N129D;V182I	4	29-01-2021
	T113I;V290F	2	04-11-2020		P43L;A100S	2	2020-10		S221A;P451S	4	08-04-2021
	T31I;M501I	2	11-10-2020		P43L;E453D	2	2020-10		V125F;V381L	4	20-12-2020
USA	N129D;P443S	3	13-01-2021		P43L;L157F	2	26-12-2020		A320V;D496Y	3	29-04-2020
	P140S;D432G	3	27-04-2020	Mexico	D291Y;R525K	2	16-01-2021		G114C;N129D	3	21-01-2021
	P451S;K469R	3	25-03-2021		D324E;L409F	2	22-04-2021		H26Y;A274S	3	06-04-2020
	A225S;K349N	2	30-01-2021		T31I;M315I	2	18-05-2021		H373Y;V381L	3	08-12-2020
	G6R;L177F	2	24-03-2021	Switzerland	M315I;A319S	10	30-12-2020		K13R;N129D	3	2021-03
	H26Y;V136F	2	15-01-2021		I15T;T524I	7	17-12-2020		N129D;A360V	3	06-02-2021
	I45T;P451S	2	06-04-2021		E204D;T219I	3	07-11-2020		N129D;D324Y	3	07-03-2021
	K349N;V421I	2	06-03-2021	USA	N129D;L383I	2	13-03-2021		N67T;N129D	2	02-02-2021
	M153I;Q254K	2	23-04-2021		N129D;P239S	2	20-10-2020		P43L;D144Y	2	09-12-2020
	M62I;M315I	2	23-02-2021		N129D;P355S	2	23-03-2021		T97I;V381L	2	05-10-2020
	M72I;V341I	2	07-04-2021		N129D;R476S	2	18-03-2021		**V263F;P412H**	2	23-03-2021
	N129D;D258N	2	26-12-2020		N129D;S450G	2	19-01-2021		V263F;V381L	2	22-03-2021
	N129D;D415G	2	16-11-2020		N129D;T131I	2	04-02-2021		V4L;N129D	2	11-12-2020

Table 8 comprehensively presented the total number of co-mutations and unique co-mutations, along with their respective percentages, organized by geographic location as well as on a global scale. This tabulated data provides insights into the distribution and prevalence of these co-mutations across different regions, offering a broader perspective on the mutation landscape within NSP14 variants.

Table 8: Frequencies and percentages of co-occurring mutations in NSP14 variants available in the various countries

Country	**Total Co-mutatons**	**Total Unique Co-mutations**	**% of Unique Co-mutations (Country-wise)**	**% of Unique Co-mutations (Globally)**
UK	42	27	64.29	12.16
USA	84	59	70.24	26.58
India	5	2	40.00	0.90
Canada	9	3	33.33	1.35
Denmark	10	5	50.00	2.25
France	9	2	22.22	0.90
Italy	3	1	33.33	0.45
Japan	18	15	83.33	6.76
Mexico	6	3	50.00	1.35
Switzerland	6	3	50.00	1.35

The analysis of Table 8 revealed intriguing insights into unique co-occurring mutations across different geographic locations. Notably, the USA exhibited the highest percentage, with 26.58% of unique co-occurring mutations. The UK and Japan followed with 12.16% and 6.76%, respectively, while other geographic locations displayed significantly lower percentages of unique co-occurring mutations. Worth mentioning is the observation that out of the 18 detected co-occurring mutations in Japan, 15 were unique to that specific geographic location. Additionally, the USA had the next highest percentage of unique co-occurring mutations (Table 8).

To further elucidate the dynamics of co-occurring mutations within SARS-CoV-2 NSP14 variants, a co-occurring mutation flow diagram was created (Figure 2). This diagram provides a visual representation of how these mutations are

interconnected. Additionally, the predicted pathogenicity of each mutation, whether neutral or deleterious, was indicated based on predictions from the PredictSNP web tool (Table 12).

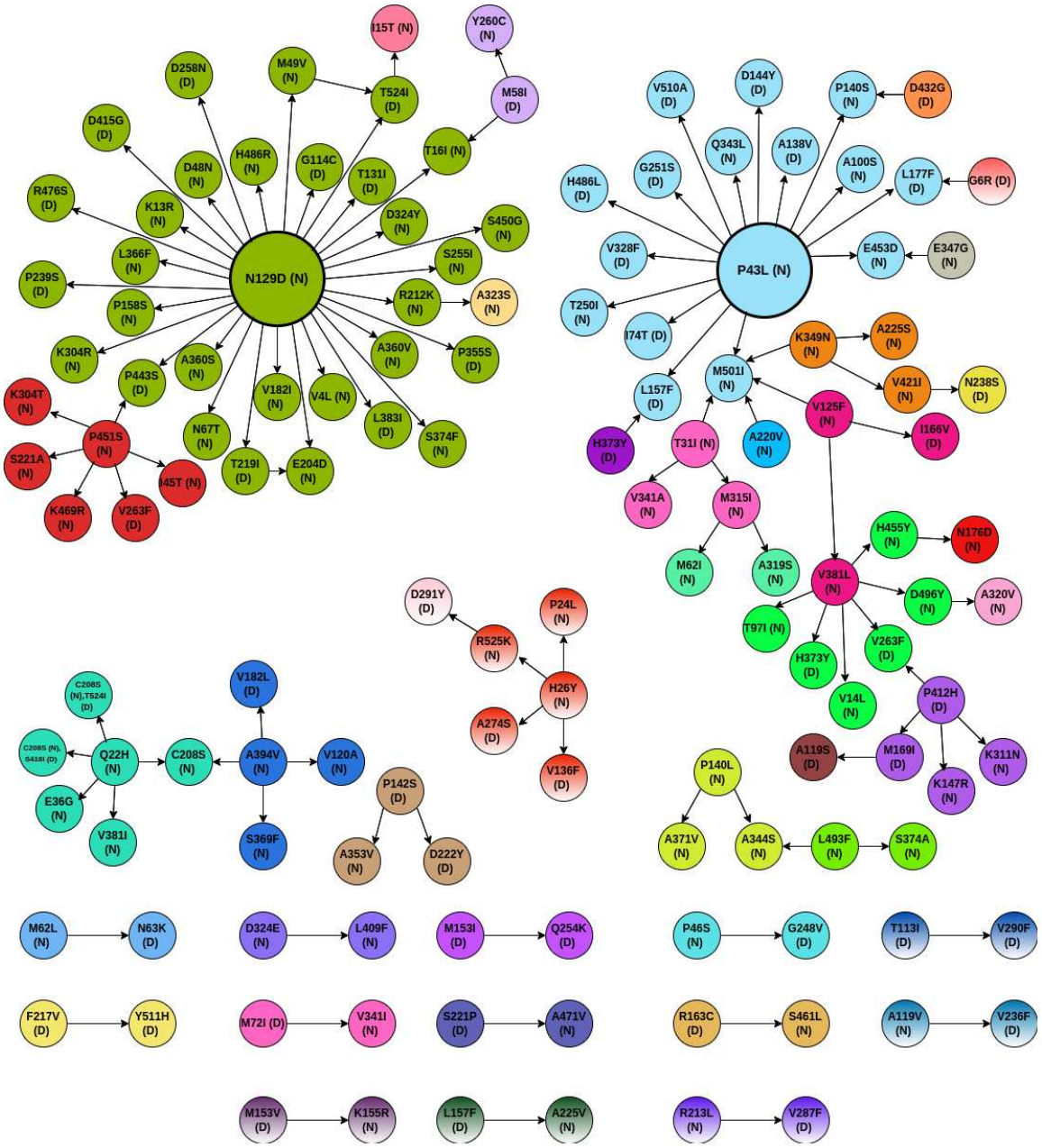

Figure 2: Co-mutation flow in SARS-CoV-2 NSP14 variants. The diagram represents the different co-mutations that are associated with each mutation. Note: 1. The colour coding has been done to understand which of the mutations are associated with a particular mutation. 2. The length of the arrow is just for representation and not indicative of anything. 3. For some mutations which are associated with more than one mutation, the colour coding of any one of the mutations is used. The **N** and **D** represent Neutral and Deleterious mutations, respectively.

Analysis of co-mutations in SARS-CoV-2 NSP14 variants revealed some interesting patterns. N129D emerged as the mutation with the highest number, with 31 co-mutations associated with it, followed by P43L with 15 co-mutations. Moreover, Q22H was identified with two triple-mutations, C208S and S418I, as well as C208S and T524I. Interestingly, both Q22H and C208S mutations were classified as neutral, while the additional co-occurring mutations, S418I and T524I, were deemed deleterious based on their pathogenicity.

Further classification of co-occurring mutations of length two was based on pathogenicity, resulting in three distinct classes: deleterious-deleterious (DD), deleterious-neutral (DN), and neutral-neutral (NN). Nine DD-type co-mutations, marked in bold in Table 7, were identified. Conversely, there were 47 DN-type and 61 NN-type co-mutations, revealing a range of mutation combinations with varying pathogenic implications (Table 9). These findings underscore the complexity of mutation interactions within NSP14 variants.

Table 9: Three different classes of co-occurring mutations in SARS-CoV-2 NSP14 variants.Yellow, blue, and green marked co-occurring mutations belong to the ExoN, N7-MTase, and both (one was from ExoN & another was from N7-MTase) domains, respectively.

D-D	DN	DN	NN	NN	NN
M153I, Q254K	R213L, V287F	P43L, A138V	D324E, L409F	V125F, M501I	N129D, P158S
T113I, V290F	A394V, V182L	P43L, L177F	Q22H, C208S	M501I, K349N	N129D, K304R
F217V, Y511H	P142S, A353V	P140S, D432G	Q22H, V381I	K349N, A225S	N129D, A360S
P142S,D222Y	H26Y, V136F	N129D, T131I	Q22H, E36G	K349N, V421I	N129D, N67T
M169I, A119S	H26Y, A274S	N129D, T524I	A394V, C208S	M315I, A319S	N129D, V182I
P412H, M169I	P412H, K147R	T524I, I15T	A394V, V120A	M315I, M62I	N129D, E204D
P412H, V263F	P412H, K311N	N129D, D258N	A394V, S369F	T31I, M315I	N129D, V4L
L157F, H373Y	V263F, V381L	N129D, D415G	H26Y, R525K	T31I, V341A	N129D, S374F
L177F, G6R	V381L, H373Y	N129D, R476S	H26Y, P24L	M501I, A220V	N129D, A360V
DN	**DN**	**DN**	**NN**	**NN**	**NN**
N129D, G114C	V125F, I166V	N129D, P239S	P140L, A371V	M501I, T31I	N129D, R212K
M62L, N63K	V421I, N238S	N129D, P443S	P140L, A344S	P43L, M501I	R212K, A323S
P46S, G248V	P43L, L157F	N129D, T219I	L493F, A344S	P43L, T250I	N129D, S255I
A119V, V236F	P43L, I74T	T219I, E204D	L493F, S374A	P43L, Q343L	N129D, S450G
R163C, S461L	P43L, V328F	N129D, L383I	V381L, V14L	P43L, P140S	N129D, D324Y
A471V, S221P	P43L, H486L	N129D, P355S	V381L, D496Y	P43L, A100S	N129D, T16I
M72I, V341I	P43L, G251S	M58I, T16I	D496Y, A320V	P43L, E453D	P451S, K304T
M153V, K155R	P43L, V510A	M58I, Y260C	V381L, T97I	E453D, E347G	P451S, S221A
L157F, A225V	P43L, D144Y	M49V, T524I	V381L, H455Y	N129D, D48N	P451S, K469R
	D291Y, R525K	P443S, P451S	H455Y, N176D	N129D, K13R	P451S, I45T
		P451S, V263F	V125F, V381L	N129D, L366F	N129D, H486R
					N129D, M49V

The tracking of co-mutated sequences in SARS-CoV-2 NSP14 variants provides valuable insights into the spread and evolution of these mutations. In Table 10, the first detected geographic location and the respective dates of the co-mutated sequences are documented. Additionally, Table 10 presents the subsequent spread of these co-mutated sequences across other geographic locations. These details shed light on the geographical and temporal dynamics of co-mutations within NSP14 variants, contributing to the understanding of their emergence and dissemination.

Table 10: Geo-location and date of first identification and subsequent detection (with 100% identical) of co-mutated sequences

CS	Origin of 1st Identification	Date of 1st Identification	Number of identical sequences	Geo-locations of the identical sequences
CS1	Japan	2020-09	2	Japan: Kanto
CS3	Japan	11-11-2020	5	Japan: Kanto
CS6	USA: Minnesota	12-08-2021	1	USA: Minnesota
CS9	USA: FL	01-12-2020	1	USA: Florida
CS10	Japan:Hokkaido	30-11-2020	1	Japan:Hokkaido
CS16	USA	25-04-2020	2	USA
CS17	USA: Florida	24-03-2021	2	USA: Florida
CS21	USA: New Jersey	20-03-2021	1	USA: New Jersey
CS28	USA: California	18-03-2021	4	USA: California
CS29	USA: California	25-05-2021	1	USA: California
CS32	USA	01-12-2020	9	USA: California, Georgia, North Carolina
CS35	USA: FL	04-02-2021	13	USA: Florida, California
CS36	USA: New York	25-03-2021	1	USA: New York
CS37	USA: Maryland	28-12-2020	6	USA: Maryland, Pennsylvania, Snohomish County
CS40	USA: Massachusetts	29-04-2020	3	USA: Massachusetts
CS41	USA: Pennsylvania	23-03-2021	1	USA: Pennsylvania
CS42	USA: New Jersey	31-03-2021	2	USA: New Jersey, New York
CS43	USA: West Virginia	16-02-2021	32	USA: West Virginia, Virginia, Minnesota, Maryland Tennessee, North Carolina, Pennsylvania, Florida
CS52	USA: Washington,SC	4-6-2020	2	USA: Washington,Snohomish County
CS59	USA: Massachusetts	4-23-2021	3	USA: New Hampshire, Massachusetts
CS61	USA: Massachusetts	12-31-2020	15	USA: Massachusetts, West Virginia, Florida
CS62	USA: CA	1-9-2021	20	USA: CA
CS63	USA: GA	12-1-2020	8	USA: West Virginia, GA, Texas, New York, Mississippi, West Virginia
CS64	USA: CA	2-4-2021	1	USA: CA
CS65	USA: CA	12-2-2020	5	USA: CA, Contra Costa County, Minnesota, Pennsylvania
CS66	USA: Alaska	11-15-2020	30	USA: Alaska, Virginia, California, North Carolina Michigan, Arizona, Georgia, Pennsylvania
CS67	USA	12-1-2020	4	USA: California,Los Angeles County
CS68	USA: TX	1-6-2021	5	USA: Texas, Pennsylvania, Alabama, Madison County, Florida, Massachusetts
CS70	USA	1-7-2021	2	USA: Virginia
CS71	USA: Virginia	9-3-2020	7	USA: Florida, Virginia, Pennsylvania, New York, Vermont
CS72	USA: Alabama, Madison County	10-27-2020	6	USA: Indiana, Michigan, Alabama, Missesota
CS73	USA	1-11-2021	5	USA: Texas
CS74	USA: Michigan	3-23-2021	1	USA: Michigan
CS76	USA: South Dakota	11-18-2020	6	USA: Nebraska, South Dakota, Minnesota
CS77	USA	8-8-2020	5	USA: Arizona, Massachusetts, Nevada
CS78	USA	12-2-2020	7	USA: North Carolina, Kansas, West Virginia, Texas, New York
CS79	USA: California	12-19-2020	33	USA: California, North Carolina, West Virginia, Minnesota, Wisconsin Alaska, Indana, Texas, Massachusetts, Illinois, New Hampshire Georgia, Missouri, Pennsylvania, South Carolina,
CS80	USA: Texas	1-23-2021	2	USA: Texas
CS81	USA: Wisconsin	3-23-2021	3	USA: Wisconsin, Tennessee
CS82	USA: MA	1-6-2021	4	USA: Massachusetts, Texas
CS83	USA: Massachusetts	3-13-2021	2	USA: Massachusetts
CS84	USA: TX	8-30-2020	21	USA: North Carolina, West Virginia, California, Missesota, Pennsylvania, Texas, Maine, Rhode Island
CS85	USA	11-18-2020	8	USA: Michigan, Texas, Michigan, Massachusetts, Wisconsin
CS86	USA: Indiana	1-24-2021	2	USA: Indiana
CS87	USA: Pennsylvania	12-10-2020	8	USA: Pennsylvania, Florida, Ohio, Minnesota
CS88	USA: Wisconsin, Dane County	12-29-2020	6	USA: Wisconsin, Florida
CS89	USA: Tennessee	3-4-2021	1	USA: Tennessee
CS90	USA: Texas	3-14-2021	13	USA: Texas, California
CS91	USA: Maryland	4-15-2021	2	USA: Maryland
CS92	USA: Michigan	4-8-2021	2	USA: Michigan
CS93	USA: MD	3-16-2021	2	USA: MD
CS94	Ghana	1-9-2021	6	Ghana, USA: Colorado, North Carolina
CS95	USA: Indiana	4-20-2021	1	USA: Indiana
CS96	USA: SC, GEORGETOWN	7-10-2020	4	USA: Virginia, Kentucky, North Carolina, SC, Georgetown
CS103	USA: Utah	6-29-2021	156	USA (Various states), Egypt
CS104	USA: Louisiana	7-31-2021	8	USA: Louisiana, Michigan, Massachusetts, Ohio, Oregon, Kentucky
CS105	USA: Colorado	7-8-2021	46	USA
CS108	New Zealand: Auckland	12-8-2020	1	New Zealand: Auckland
CS110	India: Gujarat,	6-25-2020	2	India
CS113	USA: Washington	07-04-2021	5	USA: Washington
CS120	USA: California	12-19-2020	33	USA

3.2.2. Phylogenetic variety of co-mutated NSP14 sequences

To gain a better understanding of the co-mutated sequences in Table 14, they were grouped into different clusters based on amino acid sequence homology with the reference SARS-CoV-2 NSP14 sequences. This clustering process is visualized in Figure 3, allowing for the identification of patterns and relationships among these sequences. Clustering based on sequence homology helps categorize co-mutated sequences with similar genetic characteristics, aiding in the analysis of their potential functional significance and evolutionary trends.

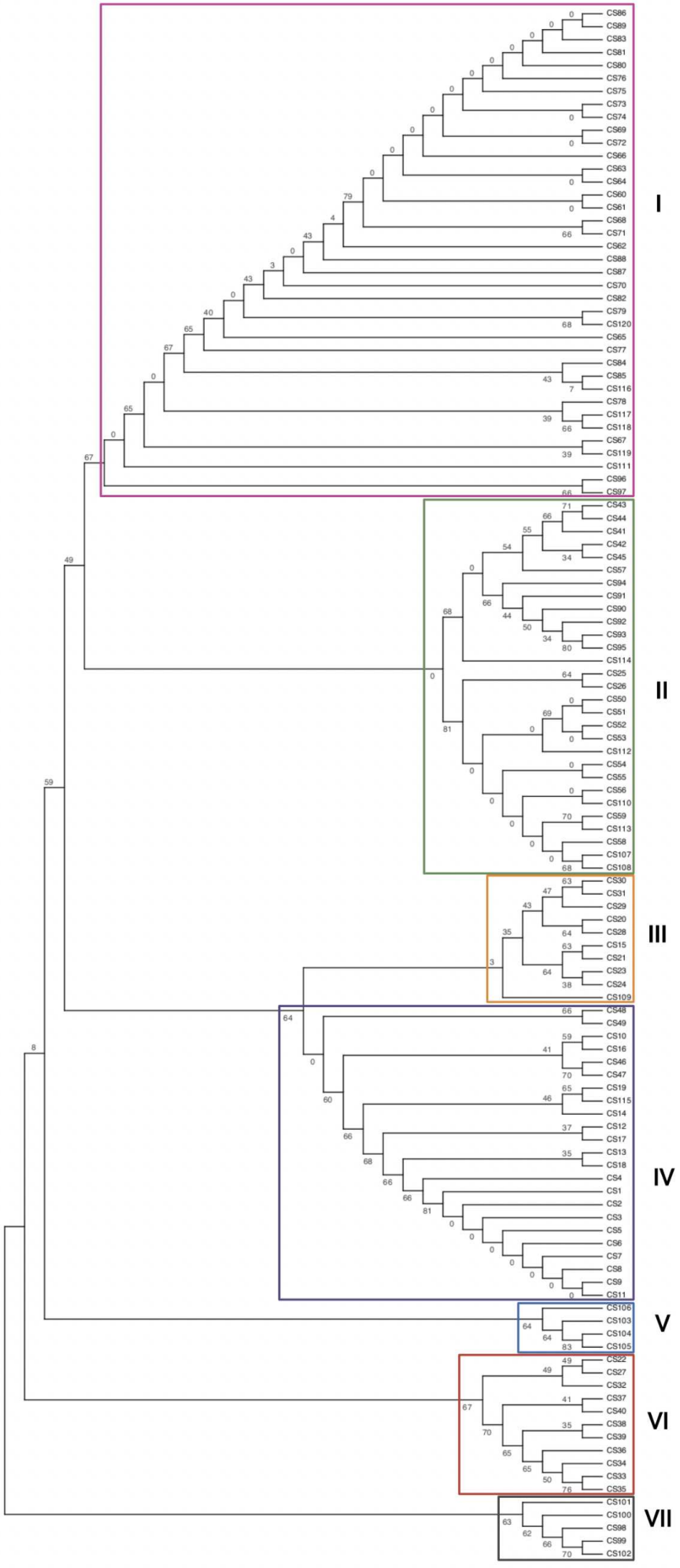

Figure 3: Phylogenetic relationship based on amino acid sequence homology of 120 co-mutated NSP14 sequences

120 co-mutated sequences CS_1 to CS_{120} were diversely allied into seven clusters. It turned out that reference SARS-CoV-2 NSP14 (YP_009725309) placed into seven distinct set of clusters as derived (and marked) in Figure 3 due to 120 co-occurring mutations. This diversity indicates wide span of varied functionalities of the NSP14 sequences.

3.3. Interacting residues of SARS-CoV-2 NSP14 sequences and mutations

A total of 146 interacting residues (marked red in Figure 4) among 527 residues present in the reference SARS-CoV-2 NSP14 sequence (YP_009725309) were predicted using the web-suite [29].

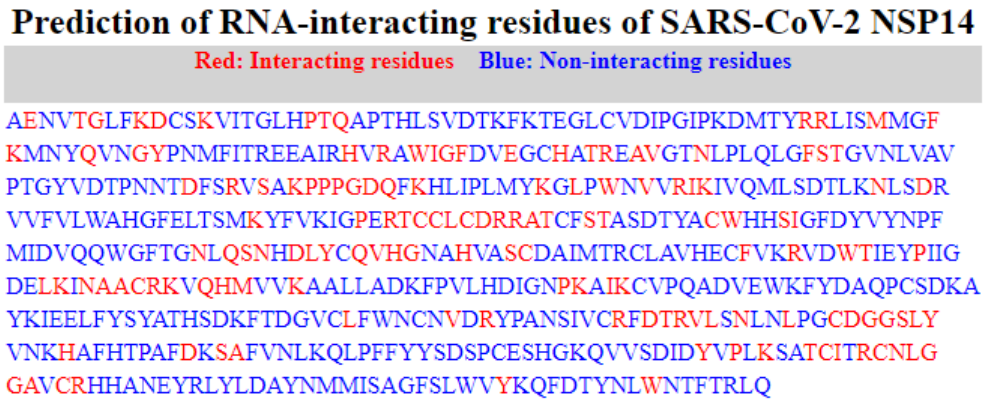

Figure 4: Interacting residues of wild-type SARS-CoV-2 NSP14 sequence

The analysis of the 146 interacting residues in Figure 4 revealed several interesting observations:

- Among these interacting residues, 113 possessed various mutations, indicating their potential role in the genetic diversity of NSP14.
- Specifically, 42 of these interacting residues were associated with only deleterious mutations, emphasizing their significance in terms of potential functional implications.
- On the other hand, 43 interacting residues had exclusively neutral mutations, suggesting that they may not directly impact the function or stability of the protein
- Additionally, a subset of 28 interacting residues exhibited mixed mutations, including both neutral and deleterious types, which could lead to complex effects on the the behaviour of the protein.
- Notably, a total of 33 interacting residues remained invariant, suggesting their crucial role in maintaining protein stability or function (Figure 5).

These findings collectively provide insights into the interplay between mutations and interacting residues within NSP14, contributing to our understanding of its structural and functional aspects.

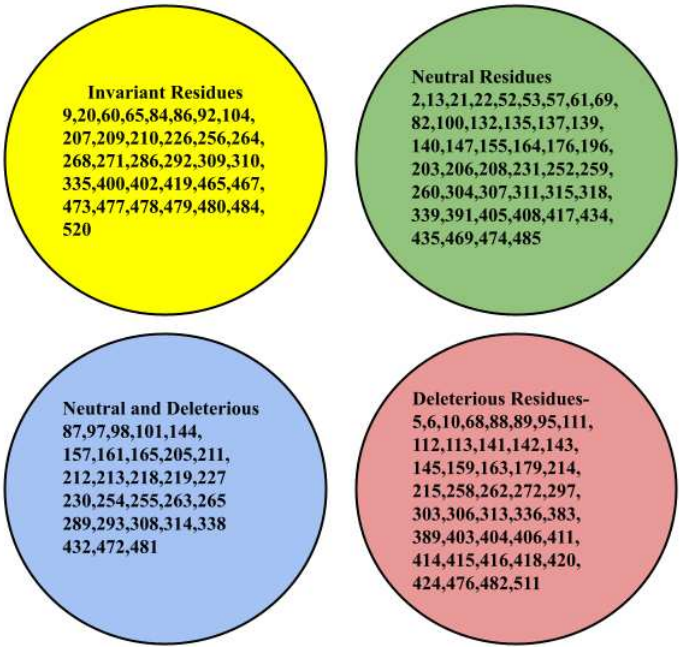

Figure 5: Three disjoint classes of interacting residues based on pathogenicity of mutations were detected. Invariant interacting residues set is marked in yellow.

3.4. Variations and clustering of unique SARS-CoV-2 NSP14 variants

3.4.1. Amino acid frequency distribution and clustering

To gain a deeper understanding of the quantitative variations among the 3953 unique SARS-CoV-2 NSP14 variants, several analyses were conducted:

A. Amino Acid Frequency Distribution: The frequency of each amino acid across all 3953 NSP14 sequences was computed and visualized, providing insights into the prevalence of specific amino acids within the variants (Figure 6(A)).

B. Pairwise Distance Calculation: Pairwise distances among all 3953 NSP14 variants were calculated, resulting in a distance matrix. This distance matrix was then plotted, revealing the genetic dissimilarity or relatedness among the variants (Figure 6(B)).

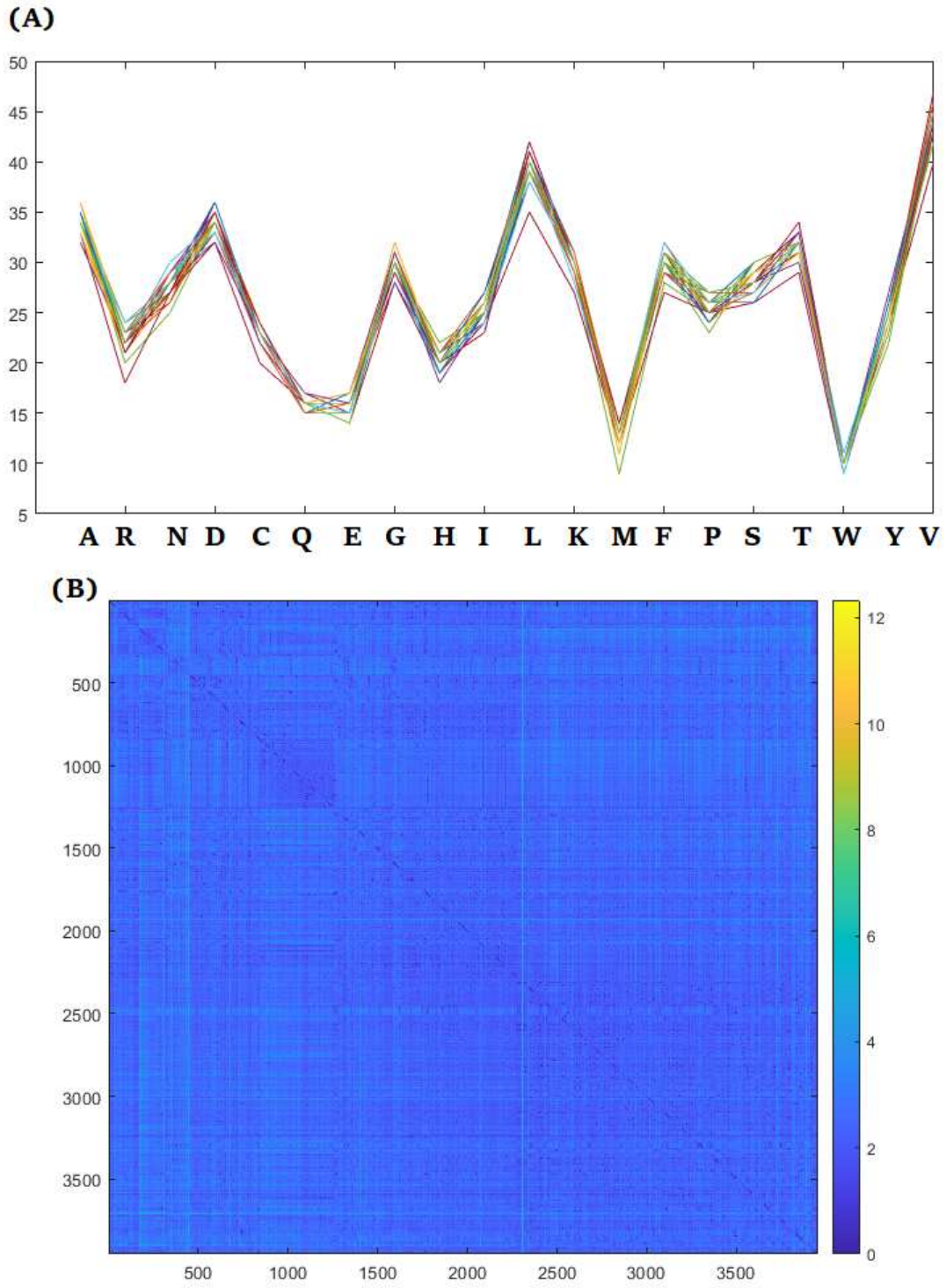

Figure 6: (A): Frequency distribution of each amino acids in 393 SARS-CoV-2 NSP14 sequences, (B): Pairwise distance among 3953 SARS-CoV-2 NSP14 sequences

The analysis of amino acid frequency distribution among the 3953 NSP14 sequences provided several interesting observations:

- Valine (V) was the most abundant amino acid, with a frequency ranging from 40 to 47 across the NSP14 variants.
- Conversely, tryptophan (W) was the least prevalent amino acid, with a presence ranging from 9 to 11 in the NSP14 variants.
- Amino acids Q and E had nearly identical frequencies, each occurring 16 times.
- Similar frequencies were also observed for the amino acid triplet (G, F, K) and the amino acid pairs (A, D), (R, C), and (P, Y).

Additionally, the clustering analysis revealed that the 3953 NSP14 sequences could be grouped into six distinct clusters: cluster:0, cluster:1, cluster:2, cluster:3, cluster:4, and cluster:5. These clusters contained varying numbers of NSP14 sequences, with cluster:0 being the largest, consisting of 2683 sequences, and cluster:5 being the smallest, with 16 sequences. A few sequences remained unclassified, denoted as cluster:-1 (Figure 7 and **Supplementary file-1**). These clusters provide insights into the genetic diversity and relatedness among the NSP14 variants.

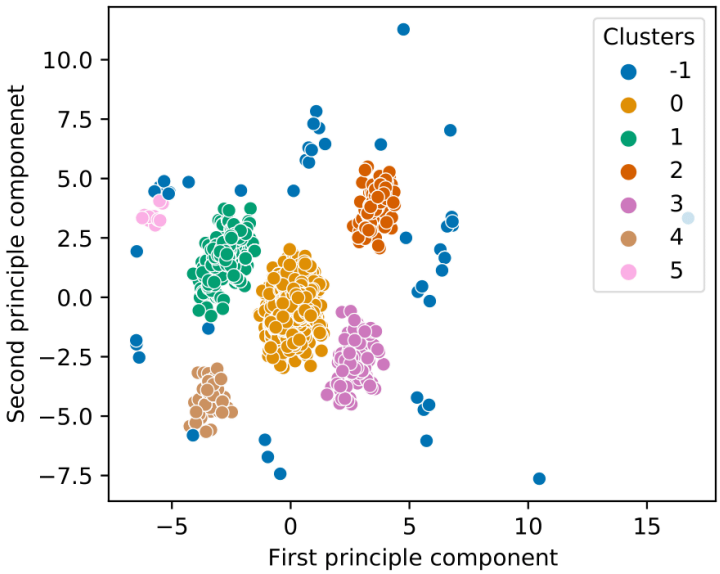

Figure 7: Clustering of NSP14 variants based on the amino acid frequency distribution of 3953 SARS-CoV-2 NSP14 sequences

The presence of 55 unclassified NSP14 sequences suggests that these sequences may not strongly align with any of the defined clusters. The existence of a dominant cluster indicates that the smaller clusters likely branched off recently from this dominant cluster, reflecting the evolutionary relationships among these variants.

It's noteworthy that the unclassified cluster is equidistant from the remaining clusters, suggesting that these unclassified sequences may represent a distinct group with genetic characteristics that differ from those in the defined clusters. This equidistance indicates a unique evolutionary trajectory for this group of NSP14 sequences. Further analysis and investigation may provide insights into the specific genetic features and evolutionary history of these unclassified sequences.

3.4.2. Quantitative physicochemical properties and clustering

The analysis of physicochemical properties for the 3953 SARS-CoV-2 NSP14 sequences provided valuable insights:

1. Isoelectric Points (PIs) Pattern: Among all the patterns of the measures plotted in Figure 8, the pattern of the isoelectric points (PIs) for the NSP14 variants exhibited a highly non-linear trend compared to the other quantitative measures obtained. This non-linearity in PIs suggests that the variations in charge distribution across the NSP14 sequences are distinct and potentially have functional implications.

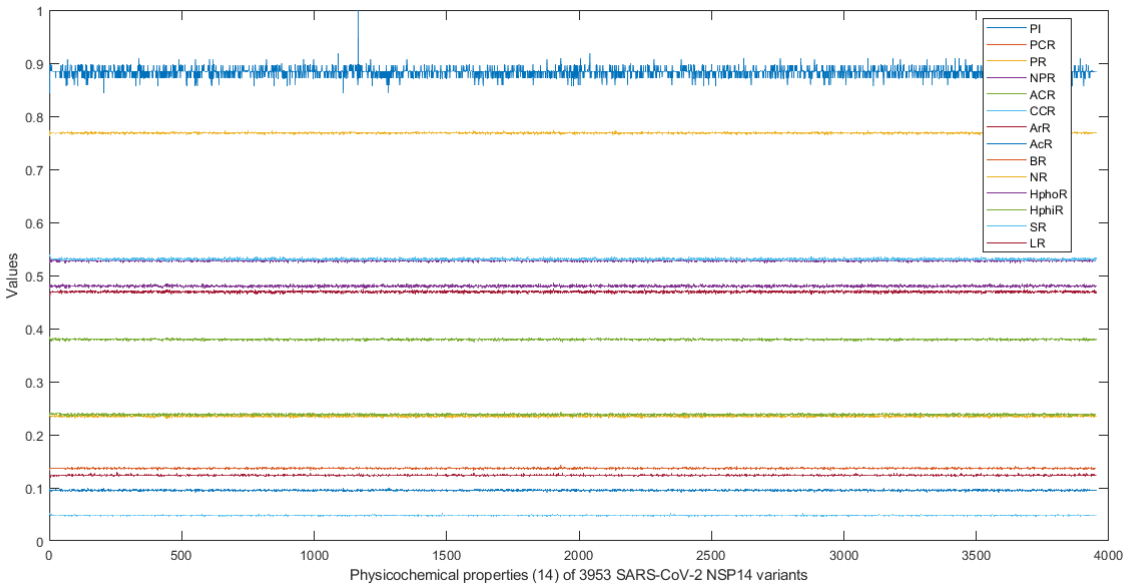

Figure 8: physicochemical features of 3953 SARS-CoV-2 NSP14 variants

2. Principal Component Analysis (PCA): Based on the quantitative features of each NSP14 sequence, a PCA analysis was conducted, resulting in the formation of three disjoint clusters (Figure 9). PCA is a powerful technique for dimensionality reduction and data visualization, allowing for the identification of underlying patterns and relationships within the dataset. These clusters provide a structured way to categorize and understand the variations in physicochemical properties among the NSP14 sequences.

These analyses contribute to the understanding of the physicochemical characteristics and variations present within the NSP14 variants, shedding light on potential functional and structural implications.

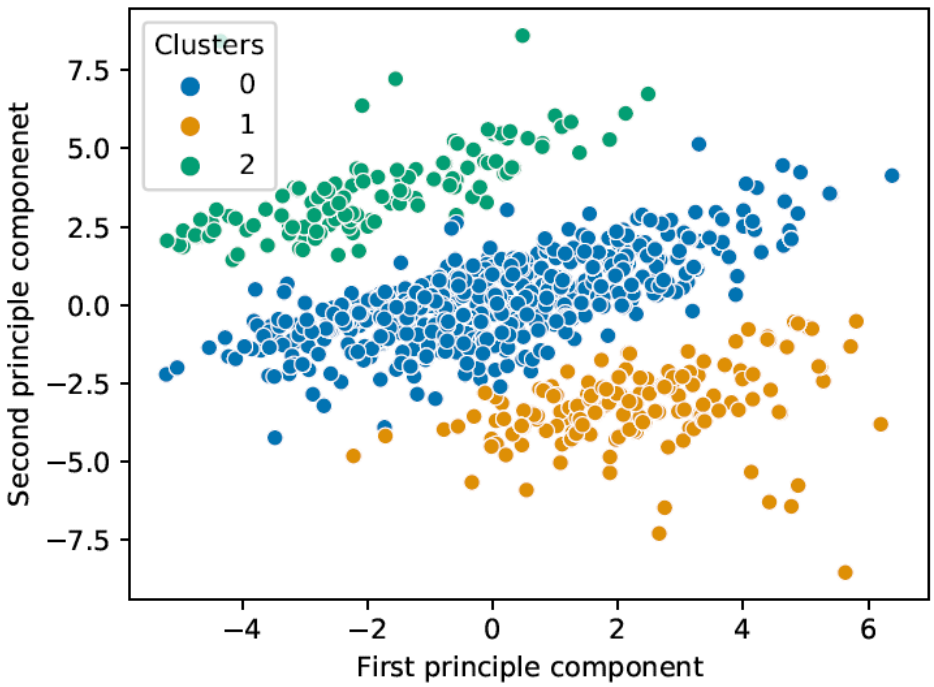

Figure 9: Clusters 0, 1, and 2 contain 3201, 297, and 455 respectively.

In the PCA analysis, the majority of the NSP14 sequences, specifically 3201 out of 3953, were grouped into one cluster, referred to as cluster-0 (Supplementary file-3). Additionally, two smaller clusters, cluster-1 and cluster-2, were identified, containing 297 and 455 NSP14 variants, respectively. This clustering provides a structured way to categorize the NSP14 sequences based on their physicochemical properties, revealing distinct subgroups within the dataset. These subgroups may have unique characteristics and functional implications worth exploring further.

4. Discussion and Concluding Remarks

The NSP14 protein in SARS-CoV-2 is a nonstructural protein consisting of 527 amino acids and contains two distinct domains: the ExoN domain at the N-terminal and the N7-MTase domain at the C-terminal. The ExoN domain is particularly noteworthy for its proofreading activity, which plays a critical role in ensuring the fidelity of viral replication. NSP14 shares sequence similarity with other coronaviruses such as SARS-CoV and MERS-CoV [31].

Certain residues within NSP14 are typically conserved, but mutations in these residues can impact viral replication. Notably, mutations like R310A and F426A can affect viral replication, while H424A results in a crippled phenotype in SARS-CoV-2, MERS-CoV, and SARS-CoV [21]. The conservation of these residues, such as R310 and F426, across different NSP14 variants underscores their potential as drug targets, as mutations in these residues can influence viral phenotype and replication.. Analyzing the NSP14 sequences from various strains of SARS-CoV-2 revealed the presence of 962 mutations in total (Figure 4). Among these, 548 were classified as neutral mutations, and 414 were considered deleterious [32]. Deleterious mutations are those that result in a loss of gene function or impaired protein production, and they play a role in down-regulating the host immune response. It is worth noting that the balance between deleterious and neutral mutations in NSP14 is crucial for viral fitness and adaptation [33].

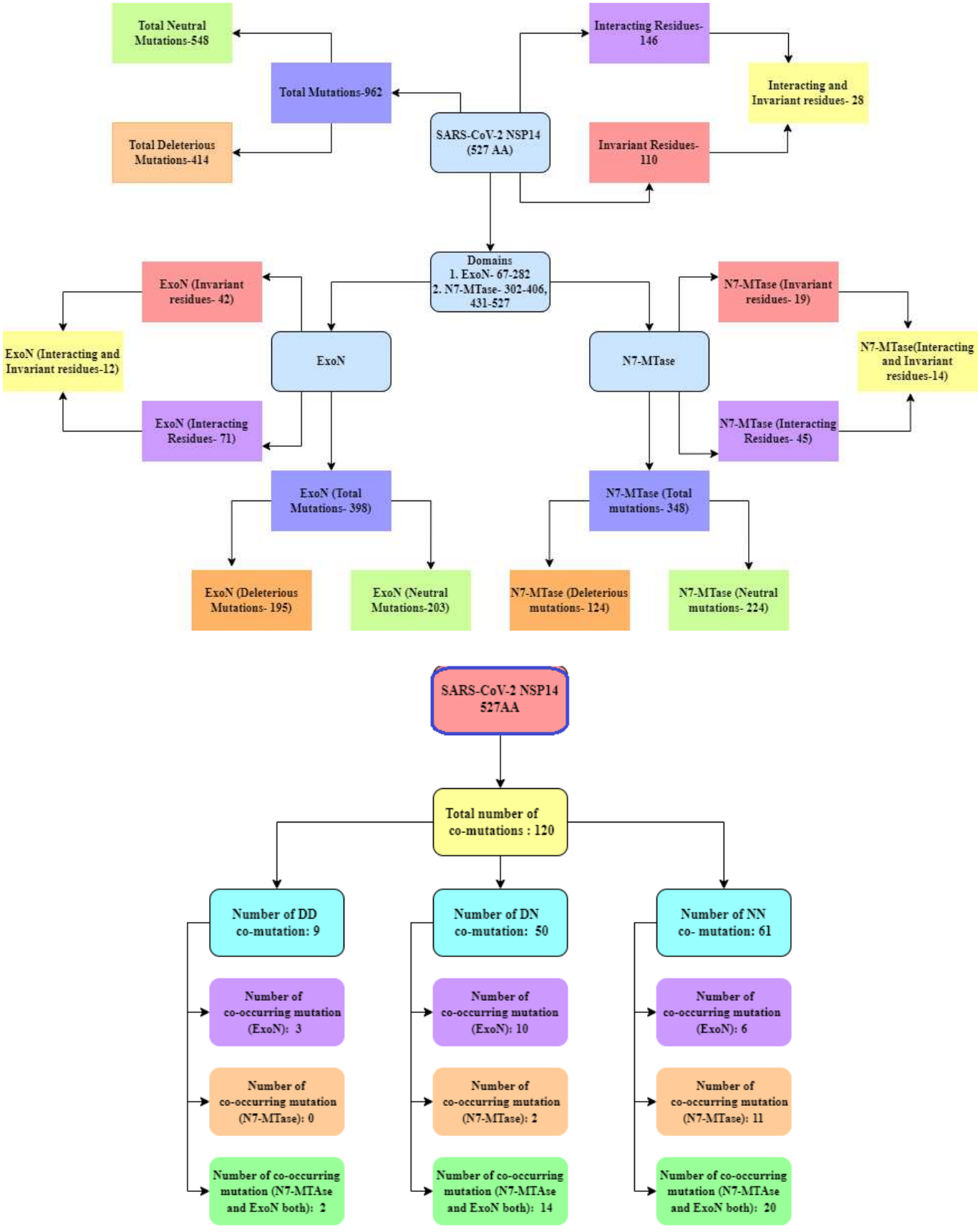

Figure 10: Schematic snapshot of frequency of mutations and co-occurring mutations, invariant residues, and interacting residues in ExoN and N7-MTase domains in SARS-CoV-2 NSP14

The ExoN domain of NSP14 exhibited a higher number of total mutations (398) compared to the N7-MTase domain (348 mutations). Inactivation of the ExoN domain can be lethal for SARS-CoV-2, and interestingly, ExoN mutations did not affect the activity of the N7-MTase domain [34]. Key residues in the ExoN region that interact with RNA are conserved across different coronaviruses, suggesting a shared RNA substrate recognition mechanism [35]. The N7-MTase domain of

NSP14 is responsible for RNA capping and modulating host gene expression changes and translational repression [36, 23]. Mutations in key residues of the N7-MTase domain can interfere with viral replication, and these residues are conserved across different coronavirus families. Our analysis identified 19 invariant residues across the N7-MTase domain, with 14 of them being both invariant and interacting, indicating their crucial role in viral replication and RNA capping [21].

Co-mutations were also analyzed, and 120 co-mutations were identified across different strains. Neutral-neutral (NN) and deleterious-neutral (DN) co-mutations were the most common types. The presence of neutral mutations in co-mutation pairs can lead to stabilizing effects, promoting the survival and fitness of co-mutated strains [37]. Co-mutations can contribute to the development of different viral strains, increasing genomic diversity and potentially affecting viral fitness [38].

Overall, understanding the genetic variations and interactions within NSP14 is essential for unraveling the mechanisms underlying viral replication, evolution, and potential drug targets. Further research is needed to validate the impact of co-mutations on viral fitness and adaptation.

Acknowledgements

We gratefully acknowledge the authors from the originating laboratories responsible for obtaining the specimens and the submitting laboratories where genetic sequence data were generated and shared via the NCBI and GISAID Initiatives, on which this research is based.

Authors' Contributions

SSH designed the study. SSH, DN, TB, PB contributed to the implementation of the research, to the analysis of the results. SSH, TB, and PB wrote the initial draft of the manuscript. All authors reviewed and edited. All authors read final version and approve.

5. Declaration of competing interest

Authors declare no competing interest to declare.

References

[1] Y. M. Bar-On, A. Flamholz, R. Phillips, R. Milo, Sars-cov-2 (covid-19) by the numbers, elife 9 (2020) e57309.

[2] P. V'kovski, A. Kratzel, S. Steiner, H. Stalder, V. Thiel, Coronavirus biology and replication: implications for sars-cov-2, Nature Reviews Microbiology 19 (3) (2021) 155–170.

[3] M. Kaur, A. Sharma, S. Kumar, G. Singh, R. P. Barnwal, Sars-cov-2: Insights into its structural intricacies and functional aspects for drug and vaccine development, International Journal of Biological Macromolecules 179 (2021) 45–60.

[4] I. Astuti, et al., Severe acute respiratory syndrome coronavirus 2 (sars-cov-2): An overview of viral structure and host response, Diabetes & Metabolic Syndrome: Clinical Research & Reviews 14 (4) (2020) 407–412.

[5] M.-Y. Wang, R. Zhao, L.-J. Gao, X.-F. Gao, D.-P. Wang, J.-M. Cao, Sars-cov-2: structure, biology, and structure-based therapeutics development, Frontiers in cellular and infection microbiology 10 (2020) 587269.

[6] M. T. Khan, M. Irfan, H. Ahsan, A. Ahmed, A. C. Kaushik, A. S. Khan, S. Chinnasamy, A. Ali, D.-Q. Wei, Structures of sars-cov-2 rna-binding proteins and therapeutic targets, Intervirology 64 (2) (2021) 55–68.

[7] S. Elbe, G. Buckland-Merrett, Data, disease and diplomacy: Gisaid's innovative contribution to global health, Global challenges 1 (1) (2017) 33–46.

[8] J. Hadfield, C. Megill, S. M. Bell, J. Huddleston, B. Potter, C. Callender, P. Sagulenko, T. Bedford, R. A. Neher, Nextstrain: real-time tracking of pathogen evolution, Bioinformatics 34 (23) (2018) 4121–4123.

[9] D. Eskier, A. Suner, G. Karakülah, Y. Oktay, Mutation density changes in sars-cov-2 are related to the pandemic stage but to a lesser extent in the dominant strain with mutations in spike and rdrp, PeerJ 8 (2020) e9703.

[10] F. Ferron, L. Subissi, A. T. Silveira De Morais, N. T. T. Le, M. Sevajol, L. Gluais, E. Decroly, C. Vonrhein, G. Bricogne, B. Canard, et al., Structural and molecular basis of mismatch correction and ribavirin excision from coronavirus rna, Proceedings of the National Academy of Sciences 115 (2) (2018) E162–E171.

[11] Y. Ma, L. Wu, N. Shaw, Y. Gao, J. Wang, Y. Sun, Z. Lou, L. Yan, R. Zhang, Z. Rao, Structural basis and functional analysis of the sars coronavirus nsp14–nsp10 complex, Proceedings of the National Academy of Sciences 112 (30) (2015) 9436–9441.

[12] F. K. Yoshimoto, A biochemical perspective of the nonstructural proteins (nsps) and the spike protein of sars cov-2, The protein journal 40 (3) (2021) 260–295.

[13] S. Lin, H. Chen, Z. Chen, F. Yang, F. Ye, Y. Zheng, J. Yang, X. Lin, H. Sun, L. Wang, et al., Crystal structure of sars-cov-2 nsp10 bound to nsp14-exon domain reveals an exoribonuclease with both structural and functional integrity, Nucleic acids research 49 (9) (2021) 5382–5392.

[14] Y. Chen, H. Cai, J. Pan, N. Xiang, P. Tien, T. Ahola, D. Guo, Functional screen reveals sars coronavirus nonstructural protein nsp14 as a novel cap n7 methyltransferase, Proceedings of the National Academy of Sciences 106 (9) (2009) 3484–3489.

[15] J. B. Case, A. W. Ashbrook, T. S. Dermody, M. R. Denison, Mutagenesis of s-adenosyl-l-methionine-binding residues in coronavirus nsp14 n7-methyltransferase demonstrates differing requirements for genome translation and resistance to innate immunity, Journal of virology 90 (16) (2016) 7248–7256.

[16] M. Bouvet, I. Imbert, L. Subissi, L. Gluais, B. Canard, E. Decroly, Rna 3'-end mismatch excision by the severe acute respiratory syndrome coronavirus nonstructural protein nsp10/nsp14 exoribonuclease complex, Proceedings of the National Academy of Sciences 109 (24) (2012) 9372–9377.

[17] G. Kokic, H. S. Hillen, D. Tegunov, C. Dienemann, F. Seitz, J. Schmitzova, L. Farnung, A. Siewert, C. Höbartner, P. Cramer, Mechanism of sars-cov-2 polymerase stalling by remdesivir, Nature communications 12 (1) (2021) 279.

[18] Y. Chen, Q. Liu, L. Zhou, Y. Zhou, H. Yan, K. Lan, Emerging sars-cov-2 variants: Why, how, and what's next?, Cell Insight 1 (3) (2022) 100029.

[19] J. Kottur, K. M. White, M. L. Rodriguez, O. Rechkoblit, R. Quintana-Feliciano, A. Nayar, A. García-Sastre, A. K. Aggarwal, Structures of sars-cov-2 n7-methyltransferase with dot1l and prmt7 inhibitors provide a platform for new antivirals, PLoS Pathogens 19 (7) (2023) e1011546.

[20] M. Kosuge, E. Furusawa-Nishii, K. Ito, Y. Saito, K. Ogasawara, Point mutation bias in sars-cov-2 variants results in increased ability to stimulate inflammatory responses, Scientific reports 10 (1) (2020) 17766.

[21] N. S. Ogando, P. El Kazzi, J. C. Zevenhoven-Dobbe, B. W. Bontes, A. Decombe, C. C. Posthuma, V. Thiel, B. Canard, F. Ferron, E. Decroly, et al., Structure–function analysis of the nsp14 n7–guanine methyltransferase reveals an essential role in betacoronavirus replication, Proceedings of the National Academy of Sciences 118 (49) (2021) e2108709118.

[22] M. Pachetti, B. Marini, F. Benedetti, F. Giudici, E. Mauro, P. Storici, C. Masciovecchio, S. Angeletti, M. Ciccozzi, R. C. Gallo, et al., Emerging sars-cov-2 mutation hot spots include a novel rna-dependent-rna polymerase variant, Journal of translational medicine 18 (2020) 1–9.

[23] J. C.-C. Hsu, M. Laurent-Rolle, J. B. Pawlak, C. B. Wilen, P. Cresswell, Translational shutdown and evasion of the innate immune response by sars-cov-2 nsp14 protein, Proceedings of the National Academy of Sciences 118 (24) (2021) e2101161118.

[24] X. Niu, F. Kong, Y. J. Hou, Q. Wang, Crucial mutation in the exoribonuclease domain of nsp14 of pedv leads to high genetic instability during viral replication, Cell & Bioscience 11 (1) (2021) 106.

[25] O. Pokharkar, H. Lakshmanan, G. V. Zyryanov, M. V. Tsurkan, Antiviral potential of antillogorgia americana and elisabethae natural products against nsp16–nsp10 complex, nsp13, and nsp14 proteins of sars-cov-2: An in silico investigation, Microbiology Research 14 (3) (2023) 993–1019.

[26] J. Bendl, J. Stourac, O. Salanda, A. Pavelka, E. D. Wieben, J. Zendulka, J. Brezovsky, J. Damborsky, Predictsnp: robust and accurate consensus classifier for prediction of disease-related mutations, PLoS computational biology 10 (1) (2014) e1003440.

[27] B. E. Pickett, E. L. Sadat, Y. Zhang, J. M. Noronha, R. B. Squires, V. Hunt, M. Liu, S. Kumar, S. Zaremba, Z. Gu, et al., Vipr: an open bioinformatics database and analysis resource for virology research, Nucleic acids research 40 (D1) (2012) D593–D598.

[28] E. Capriotti, P. Fariselli, Phd-snpg: a webserver and lightweight tool for scoring single nucleotide variants, Nucleic acids research 45 (W1) (2017) W247–W252.

[29] M. Kumar, M. M. Gromiha, G. P. S. Raghava, Prediction of rna binding sites in a protein using svm and pssm profile, Proteins: Structure, Function, and Bioinformatics 71 (1) (2008) 189–194.

[30] M. Ester, H.-P. Kriegel, J. Sander, X. Xu, et al., A density-based algorithm for discovering clusters in large spatial databases with noise, in: kdd, Vol. 96, 1996, pp. 226–231.

[31] M. Tahir, Coronavirus genomic nsp14-exon, structure, role, mechanism, and potential application as a drug target, Journal of Medical Virology 93 (7) (2021) 4258–4264.

[32] M. A. Islam, S. Shahi, A. A. Marzan, M. R. Amin, M. N. Hasan, M. N. Hoque, A. Ghosh, A. Barua, A. Khan, K. Dhama, et al., Variant-specific deleterious mutations in the sars-cov-2 genome reveal immune responses and potentials for prophylactic vaccine development, Frontiers in Pharmacology 14 (2023) 1090717.

[33] S. D. Frost, B. R. Magalis, S. L. Kosakovsky Pond, Neutral theory and rapidly evolving viral pathogens, Molecular Biology and Evolution 35 (6) (2018) 1348–1354.

[34] N. S. Ogando, J. C. Zevenhoven-Dobbe, Y. van der Meer, P. J. Bredenbeek, C. C. Posthuma, E. J. Snijder, The enzymatic activity of the nsp14 exoribonuclease is critical for replication of mers-cov and sars-cov-2, Journal of virology 94 (23) (2020) 10–1128.

[35] C. Liu, W. Shi, S. T. Becker, D. G. Schatz, B. Liu, Y. Yang, Structural basis of mismatch recognition by a sars-cov-2 proofreading enzyme, Science 373 (6559) (2021) 1142–1146.

[36] M. Zaffagni, J. M. Harris, I. L. Patop, N. R. Pamudurti, S. Nguyen, S. Kadener, Sars-cov-2 nsp14 mediates the effects of viral infection on the host cell transcriptome, Elife 11 (2022) e71945.

[37] N. Periwal, S. B. Rathod, S. Sarma, G. S. Johar, A. Jain, R. P. Barnwal, K. R. Srivastava, B. Kaur, P. Arora, V. Sood, Time series analysis of sars-cov-2 genomes and correlations among highly prevalent mutations, Microbiology Spectrum 10 (5) (2022) e01219–22.

[38] D. Eskier, A. Suner, Y. Oktay, G. Karakülah, Mutations of sars-cov-2 nsp14 exhibit strong association with increased genome-wide mutation load, PeerJ 8 (2020) e10181.

Appendix

Table 11: Pathogenicity (Predicted effect) of single mutations detected in the 3953 NSP14 variants

Mutation	PredictSNP	Type	Mutation	PredictSNP	Type	Mutation	PredictSNP	Type	Mutation	PredictSNP	Type
A1S	74%	Neutral	P140L	74%	Neutral	Y260F	83%	Neutral	D379A	87%	Deleterious
A1T	60%	Neutral	P140S	83%	Neutral	Y260H	83%	Neutral	D379E	55%	Deleterious
A1V	63%	Neutral	P141S	79%	Deleterious	Y260N	60%	Neutral	D379G	87%	Deleterious
E2D	75%	Neutral	P142H	79%	Deleterious	Q262R	51%	Deleterious	D379N	60%	Neutral
N3S	75%	Neutral	P142L	87%	Deleterious	V263A	79%	Deleterious	D379Y	87%	Deleterious
V4A	63%	Neutral	P142S	79%	Deleterious	V263F	61%	Deleterious	G380C	87%	Deleterious
V4I	83%	Neutral	P142T	87%	Deleterious	V263I	83%	Neutral	V381I	75%	Neutral
V4L	83%	Neutral	G143R	87%	Deleterious	G265C	65%	Deleterious	V381L	83%	Neutral
T5A	72%	Deleterious	D144A	61%	Deleterious	G265D	64%	Deleterious	C382F	87%	Deleterious
T5I	76%	Deleterious	D144E	83%	Neutral	G265S	83%	Neutral	C382S	60%	Neutral
G6R	76%	Deleterious	D144G	61%	Deleterious	G265V	55%	Deleterious	L383I	72%	Deleterious
G6V	76%	Deleterious	D144N	61%	Deleterious	N266D	60%	Neutral	F384L	87%	Deleterious
L7F	76%	Deleterious	D144Y	76%	Deleterious	N266S	55%	Deleterious	N388D	87%	Deleterious
D10Y	72%	Deleterious	Q145H	76%	Deleterious	N266T	72%	Deleterious	V389A	87%	Deleterious
K13N	60%	Neutral	Q145K	55%	Deleterious	A267S	51%	Deleterious	V389I	76%	Deleterious
K13R	83%	Neutral	Q145L	76%	Deleterious	A267V	79%	Deleterious	D390A	87%	Deleterious
V14I	83%	Neutral	Q145P	87%	Deleterious	V269I	61%	Deleterious	D390G	72%	Deleterious
V14L	83%	Neutral	Q145R	72%	Deleterious	C272Y	87%	Deleterious	D390N	65%	Deleterious
I15S	83%	Neutral	K147R	83%	Neutral	A274S	87%	Deleterious	R391K	83%	Neutral
I15T	83%	Neutral	H148Q	76%	Deleterious	I275T	60%	Neutral	R391S	83%	Neutral
I15V	83%	Neutral	H148Y	65%	Deleterious	I275V	83%	Neutral	P393L	87%	Deleterious
T16A	83%	Neutral	L149F	72%	Deleterious	M276I	87%	Deleterious	P393S	87%	Deleterious
T16I	83%	Neutral	I150L	83%	Neutral	T277A	87%	Deleterious	A394S	83%	Neutral
T16N	83%	Neutral	I150T	61%	Deleterious	R278K	60%	Neutral	A394V	60%	Neutral
G17R	76%	Deleterious	I150V	83%	Neutral	A281G	87%	Deleterious	N395S	52%	Deleterious
G17W	76%	Deleterious	P151L	76%	Deleterious	A281S	87%	Deleterious	S396A	83%	Neutral
T21A	83%	Neutral	P151Q	60%	Neutral	A281T	87%	Deleterious	I397V	83%	Neutral
Q22H	83%	Neutral	P151S	60%	Neutral	A281V	87%	Deleterious	V398A	87%	Deleterious
Q22L	63%	Neutral	P151T	65%	Deleterious	V282A	61%	Deleterious	V398I	76%	Deleterious
A23S	65%	Deleterious	L152F	60%	Neutral	V282I	83%	Neutral	C399F	87%	Deleterious
P24L	83%	Neutral	L152I	52%	Deleterious	H283N	74%	Neutral	T403N	87%	Deleterious
P24S	65%	Neutral	M153I	51%	Deleterious	H283Q	55%	Deleterious	R404K	61%	Deleterious
P24T	83%	Neutral	M153V	61%	Deleterious	H283R	87%	Deleterious	V405L	83%	Neutral
H26N	61%	Neutral	Y154C	61%	Deleterious	H283Y	83%	Neutral	L406I	55%	Deleterious
H26Q	60%	Neutral	Y154H	83%	Neutral	E284D	83%	Neutral	S407A	61%	Deleterious
H26R	63%	Neutral	K155R	83%	Neutral	E284G	61%	Deleterious	S407F	87%	Deleterious
H26Y	75%	Neutral	G156V	87%	Deleterious	E284K	61%	Deleterious	N408D	83%	Neutral
L27F	63%	Neutral	L157F	55%	Deleterious	V287A	60%	Neutral	N408K	83%	Neutral
L27I	83%	Neutral	L157I	60%	Neutral	V287F	76%	Deleterious	N408S	83%	Neutral
L27R	72%	Deleterious	L157P	61%	Deleterious	K288E	74%	Neutral	L409F	83%	Neutral
S28C	72%	Deleterious	P158H	76%	Deleterious	K288N	68%	Neutral	L411F	76%	Deleterious
S28G	68%	Neutral	P158L	87%	Deleterious	K288R	60%	Neutral	L411M	79%	Deleterious
S28I	87%	Deleterious	P158S	74%	Neutral	R289C	65%	Deleterious	L411V	76%	Deleterious
S28N	61%	Deleterious	P158T	60%	Neutral	R289H	83%	Neutral	P412H	61%	Deleterious
V29A	55%	Deleterious	W159C	87%	Deleterious	R289L	60%	Neutral	P412L	60%	Neutral
D30E	63%	Neutral	N160H	83%	Neutral	R289S	83%	Neutral	P412S	60%	Neutral
D30G	83%	Neutral	N160S	83%	Neutral	V290F	87%	Deleterious	P412T	60%	Neutral
T31I	74%	Neutral	V161A	61%	Deleterious	V290I	63%	Neutral	G413S	87%	Deleterious
T31S	83%	Neutral	V161I	83%	Neutral	D291A	72%	Deleterious	C414F	87%	Deleterious
K32E	83%	Neutral	V161L	60%	Neutral	D291E	63%	Neutral	C414S	72%	Deleterious
K32R	83%	Neutral	V162L	65%	Deleterious	D291G	76%	Deleterious	D415A	61%	Deleterious
F33L	72%	Deleterious	R163C	87%	Deleterious	D291N	83%	Neutral	D415G	61%	Deleterious
K34E	76%	Deleterious	R163H	87%	Deleterious	D291Y	87%	Deleterious	D415Y	76%	Deleterious
K34I	87%	Deleterious	R163L	87%	Deleterious	T293A	65%	Neutral	G416S	87%	Deleterious
K34R	74%	Neutral	I164L	83%	Neutral	T293I	51%	Deleterious	G417S	60%	Neutral
T35A	75%	Neutral	I164M	75%	Neutral	T293N	83%	Neutral	S418C	79%	Deleterious
T35I	63%	Neutral	I164T	83%	Neutral	I294M	61%	Deleterious	S418G	76%	Deleterious
E36A	83%	Neutral	I164V	83%	Neutral	I294S	76%	Deleterious	S418I	87%	Deleterious
E36D	83%	Neutral	K165N	61%	Deleterious	I294T	61%	Deleterious	S418V	87%	Deleterious
E36G	83%	Neutral	K165R	83%	Neutral	I294V	75%	Neutral	Y420F	76%	Deleterious
E36V	61%	Deleterious	I166L	61%	Deleterious	E295G	55%	Deleterious	Y420S	87%	Deleterious
G37C	61%	Deleterious	I166S	87%	Deleterious	E295K	60%	Neutral	V421A	87%	Deleterious
G37S	83%	Neutral	I166V	61%	Deleterious	E295V	65%	Neutral	V421E	87%	Deleterious
G37V	61%	Deleterious	V167A	55%	Deleterious	Y296H	87%	Deleterious	V421I	63%	Neutral
C39F	76%	Deleterious	V167I	61%	Deleterious	P297H	87%	Deleterious	V421L	87%	Deleterious
V40A	74%	Neutral	V167L	87%	Deleterious	P297L	87%	Deleterious	K423R	61%	Deleterious
V40F	61%	Deleterious	Q168K	76%	Deleterious	P297S	87%	Deleterious	H424R	87%	Deleterious
D41E	74%	Neutral	M169I	61%	Deleterious	I298T	83%	Neutral	A425S	87%	Deleterious
D41N	83%	Neutral	M169T	87%	Deleterious	I298V	83%	Neutral	A425T	87%	Deleterious
D41Y	83%	Neutral	M169V	76%	Deleterious	I299M	87%	Deleterious	A425V	79%	Deleterious
I42L	75%	Neutral	S171G	74%	Neutral	I299V	51%	Deleterious	H427L	75%	Neutral
I42M	72%	Deleterious	T173A	75%	Neutral	G300C	55%	Deleterious	H427Y	61%	Deleterious
I42T	65%	Deleterious	T173I	75%	Neutral	G300D	74%	Neutral	T428A	72%	Deleterious
I42V	83%	Neutral	T173K	63%	Neutral	G300S	83%	Neutral	P429L	72%	Deleterious
P43H	52%	Deleterious	L174F	65%	Neutral	D301A	61%	Deleterious	P429S	60%	Neutral
P43L	74%	Neutral	K175E	83%	Neutral	D301E	83%	Neutral	P429T	65%	Deleterious
P43S	83%	Neutral	K175R	83%	Neutral	D301G	61%	Deleterious	A430S	60%	Neutral
G44C	60%	Neutral	N176D	83%	Neutral	D301N	83%	Neutral	A430T	61%	Deleterious
G44S	60%	Neutral	N176H	83%	Neutral	D301Y	87%	Deleterious	A430V	61%	Deleterious

Table 12: Pathogenicity (Predicted effect) of single mutations detected in the 3953 NSP14 variants (contd.)

Mutation	PredictSNP	Type	Mutation	PredictSNP	Type	Mutation	PredictSNP	Type	Mutation	PredictSNP	Type
G44V	83%	Neutral	N176S	83%	Neutral	L303P	55%	Deleterious	F431L	74%	Neutral
I45L	83%	Neutral	N176T	74%	Neutral	K304M	83%	Neutral	F431V	76%	Deleterious
I45M	60%	Neutral	L177F	55%	Deleterious	K304N	83%	Neutral	D432E	60%	Neutral
I45T	83%	Neutral	L177I	83%	Neutral	K304R	83%	Neutral	D432G	72%	Deleterious
I45V	83%	Neutral	L177V	83%	Neutral	K304T	75%	Neutral	D432N	83%	Neutral
P46F	62%	Deleterious	S178A	63%	Neutral	I305L	74%	Neutral	D432Y	76%	Deleterious
P46L	74%	Neutral	S178C	76%	Deleterious	I305T	76%	Deleterious	K433N	61%	Deleterious
P46S	83%	Neutral	D179A	87%	Deleterious	I305V	83%	Neutral	K433R	83%	Neutral
K47N	83%	Neutral	D179G	87%	Deleterious	N306K	87%	Deleterious	S434G	83%	Neutral
K47R	68%	Neutral	D179N	87%	Deleterious	N306S	76%	Deleterious	S434I	83%	Neutral
D48G	83%	Neutral	R180S	83%	Neutral	A307P	60%	Neutral	S434N	83%	Neutral
D48N	83%	Neutral	V181A	76%	Deleterious	A307S	83%	Neutral	A435S	63%	Neutral
D48Y	55%	Deleterious	V181I	83%	Neutral	A307V	65%	Neutral	A435V	60%	Neutral
M49I	83%	Neutral	V182A	61%	Deleterious	A308S	83%	Neutral	F436I	87%	Deleterious
M49L	74%	Neutral	V182I	83%	Neutral	A308T	65%	Deleterious	F436L	65%	Deleterious
M49T	74%	Neutral	V182L	61%	Deleterious	A308V	65%	Deleterious	V437A	83%	Neutral
M49V	83%	Neutral	F183L	60%	Neutral	K311E	74%	Neutral	V437F	65%	Neutral
T50A	60%	Neutral	V184F	87%	Deleterious	K311N	83%	Neutral	V437I	83%	Neutral
T50I	76%	Deleterious	V184I	60%	Neutral	K311R	83%	Neutral	N438S	60%	Neutral
T50S	83%	Neutral	L185S	63%	Neutral	K311T	83%	Neutral	K440R	75%	Neutral
Y51C	87%	Neutral	A187S	63%	Neutral	V312F	87%	Deleterious	Q441L	83%	Neutral
R52K	83%	Neutral	A187T	87%	Deleterious	V312I	60%	Neutral	L442S	51%	Deleterious
R53K	60%	Neutral	A187V	87%	Deleterious	Q313E	65%	Deleterious	L442V	60%	Neutral
L54I	55%	Deleterious	H188Q	55%	Deleterious	H314P	61%	Deleterious	P443L	76%	Deleterious
I55T	72%	Deleterious	H188R	76%	Deleterious	H314Q	74%	Neutral	P443S	87%	Deleterious
I55V	74%	Neutral	H188Y	72%	Deleterious	H314Y	60%	Neutral	P443T	87%	Deleterious
M57I	65%	Neutral	G189C	76%	Deleterious	M315I	74%	Neutral	F444Y	87%	Deleterious
M57K	74%	Neutral	F190L	83%	Neutral	M315L	83%	Neutral	Y447H	87%	Deleterious
M57T	75%	Neutral	E191D	61%	Deleterious	M315T	74%	Neutral	S448A	61%	Deleterious
M57V	65%	Neutral	T193A	60%	Neutral	M315V	83%	Neutral	D449E	75%	Neutral
M58I	72%	Deleterious	S194A	63%	Neutral	V316F	83%	Neutral	D449G	76%	Deleterious
M58L	74%	Neutral	S194C	83%	Neutral	V316I	83%	Neutral	D449N	55%	Deleterious
G59C	87%	Deleterious	M195I	72%	Deleterious	V317A	60%	Neutral	S450G	83%	Neutral
G59S	87%	Deleterious	M195T	76%	Deleterious	V317I	68%	Neutral	S450I	79%	Deleterious
G59V	87%	Deleterious	K196Q	65%	Neutral	K318R	83%	Neutral	S450N	72%	Deleterious
K61R	75%	Neutral	K196R	83%	Neutral	A319S	83%	Neutral	S450R	51%	Deleterious
M62I	83%	Neutral	Y197C	87%	Deleterious	A319T	87%	Deleterious	P451L	63%	Neutral
M62L	83%	Neutral	F198C	87%	Deleterious	A319V	51%	Deleterious	P451Q	60%	Neutral
M62T	52%	Deleterious	F198L	87%	Deleterious	A320S	72%	Deleterious	P451R	72%	Deleterious
M62V	74%	Neutral	V199A	74%	Neutral	A320T	72%	Deleterious	P451S	68%	Neutral
N63K	61%	Deleterious	V199L	72%	Deleterious	A320V	68%	Neutral	P451T	68%	Neutral
N63S	83%	Neutral	K200R	55%	Deleterious	L321V	74%	Neutral	E453A	63%	Neutral
Y64C	65%	Deleterious	I201L	74%	Neutral	L322S	74%	Neutral	E453D	83%	Neutral
N67K	74%	Neutral	I201M	60%	Neutral	L322V	75%	Neutral	E453G	52%	Deleterious
N67S	74%	Neutral	I201T	76%	Deleterious	A323S	83%	Neutral	H455Q	83%	Neutral
N67T	68%	Neutral	I201V	74%	Neutral	A323V	83%	Neutral	H455R	83%	Neutral
G68C	87%	Deleterious	P203H	60%	Neutral	D324E	71%	Neutral	H455S	83%	Neutral
Y69H	60%	Neutral	P203L	74%	Neutral	D324G	83%	Neutral	H455Y	74%	Neutral
P70L	63%	Neutral	P203S	74%	Neutral	D324H	65%	Neutral	K457R	75%	Neutral
P70S	83%	Neutral	P203T	83%	Neutral	D324N	83%	Neutral	V459I	74%	Neutral
P70T	74%	Neutral	E204A	55%	Deleterious	D324Y	83%	Neutral	V459L	83%	Neutral
N71T	83%	Neutral	E204D	83%	Neutral	K325R	83%	Neutral	V460L	83%	Neutral
M72I	51%	Deleterious	E204G	72%	Deleterious	F326I	83%	Neutral	S461A	74%	Neutral
M72L	83%	Neutral	E204Q	74%	Neutral	F326L	83%	Neutral	S461L	60%	Neutral
M72V	51%	Deleterious	E204V	63%	Neutral	F326S	83%	Neutral	D462E	83%	Neutral
I74S	87%	Deleterious	R205C	61%	Deleterious	F326Y	83%	Neutral	I463V	83%	Neutral
I74T	87%	Deleterious	R205H	60%	Neutral	P327L	74%	Neutral	D464G	76%	Deleterious
I74V	60%	Neutral	R205L	83%	Neutral	P327Q	83%	Neutral	D464N	83%	Deleterious
T75A	87%	Deleterious	T206I	83%	Neutral	P327S	83%	Neutral	D464Y	76%	Neutral
T75I	87%	Deleterious	T206S	83%	Neutral	P327T	83%	Neutral	V466L	87%	Neutral
R76C	87%	Deleterious	C208S	83%	Neutral	V328F	61%	Deleterious	K469R	83%	Neutral
R76H	87%	Deleterious	D211A	65%	Neutral	V328I	83%	Neutral	A471T	74%	Neutral
R76L	87%	Deleterious	D211G	83%	Neutral	L329F	55%	Deleterious	A471V	60%	Neutral
R76S	87%	Deleterious	D211Y	65%	Deleterious	L329I	83%	Neutral	T472A	75%	Neutral
E77D	83%	Neutral	R212G	60%	Neutral	H330R	72%	Deleterious	T472M	65%	Deleterious
E77G	51%	Deleterious	R212I	61%	Deleterious	H330Y	83%	Neutral	I474V	71%	Neutral
E78A	83%	Neutral	R212K	83%	Neutral	D331Y	87%	Deleterious	T475A	79%	Deleterious
E78K	60%	Neutral	R213C	87%	Deleterious	I332L	61%	Deleterious	R476C	87%	Deleterious
A79S	61%	Deleterious	R213H	61%	Deleterious	I332V	65%	Neutral	R476S	87%	Deleterious
A79T	87%	Deleterious	R213L	60%	Neutral	K336R	72%	Deleterious	G481C	87%	Deleterious
I80K	75%	Neutral	R213S	74%	Neutral	I338L	74%	Neutral	G481S	63%	Neutral
I80T	61%	Deleterious	A214T	76%	Deleterious	I338T	87%	Deleterious	G481V	87%	Deleterious
I80V	75%	Neutral	A214V	76%	Deleterious	I338V	75%	Neutral	A482V	87%	Deleterious
R81K	83%	Neutral	T215A	61%	Deleterious	K339N	60%	Neutral	R485G	74%	Neutral
H82Q	83%	Neutral	T215I	61%	Deleterious	K339R	83%	Neutral	R485K	83%	Neutral
H82R	83%	Neutral	C216F	75%	Neutral	V341A	83%	Neutral	H486L	51%	Deleterious
H82Y	83%	Neutral	C216L	75%	Neutral	V341I	83%	Neutral	H486R	83%	Neutral
V83A	61%	Deleterious	C216S	61%	Deleterious	V341L	60%	Neutral	H486Y	65%	Deleterious
V83I	76%	Deleterious	F217L	51%	Deleterious	P342H	83%	Neutral	A488D	87%	Deleterious

Table 13: List of single point mutations and their pathogenicity

Mutation	PredictSNP	Type	Mutation	PredictSNP	Type	Mutation	PredictSNP	Type	Mutation	PredictSNP	Type
V83L	72%	Deleterious	F217S	55%	Deleterious	P342L	83%	Neutral	A488V	51%	Deleterious
A85S	83%	Neutral	F217V	72%	Deleterious	P342S	83%	Neutral	N489D	83%	Neutral
I87L	83%	Neutral	F217Y	83%	Neutral	P342T	83%	Neutral	N489K	83%	Neutral
I87M	72%	Deleterious	S218A	60%	Neutral	Q343H	83%	Neutral	N489S	83%	Neutral
I87T	72%	Deleterious	S218F	76%	Deleterious	Q343K	83%	Neutral	E490A	60%	Neutral
I87V	83%	Neutral	T219A	51%	Deleterious	Q343L	74%	Neutral	E490D	74%	Neutral
G88S	87%	Deleterious	T219I	51%	Deleterious	Q343R	83%	Neutral	E490G	76%	Deleterious
F89C	87%	Deleterious	T219N	75%	Neutral	A344S	83%	Neutral	Y491C	87%	Deleterious
F89L	65%	Deleterious	T219S	83%	Neutral	A344T	75%	Neutral	Y491H	87%	Deleterious
V91A	63%	Neutral	A220S	83%	Neutral	A344V	74%	Neutral	L493F	75%	Neutral
V91I	61%	Deleterious	A220T	83%	Neutral	D345A	51%	Deleterious	L493S	83%	Neutral
H95L	87%	Deleterious	A220V	83%	Neutral	D345E	63%	Neutral	L493V	74%	Neutral
H95Q	87%	Deleterious	S221A	74%	Neutral	D345G	83%	Neutral	L493W	51%	Deleterious
H95Y	87%	Deleterious	S221L	83%	Neutral	D345N	83%	Neutral	Y494H	87%	Deleterious
A96S	87%	Deleterious	S221P	52%	Deleterious	D345Y	63%	Neutral	L495I	83%	Neutral
A96V	74%	Neutral	S221T	83%	Neutral	V346A	83%	Neutral	D496A	74%	Neutral
T97I	74%	Neutral	D222G	83%	Neutral	V346L	83%	Neutral	D496E	83%	Neutral
T97N	72%	Deleterious	D222N	83%	Neutral	E347A	74%	Neutral	D496G	68%	Neutral
R98G	83%	Neutral	D222V	63%	Neutral	E347D	83%	Neutral	D496H	83%	Neutral
R98I	72%	Deleterious	D222Y	52%	Deleterious	E347G	83%	Neutral	D496N	83%	Neutral
R98K	83%	Neutral	T223A	83%	Neutral	E347K	83%	Neutral	D496Y	63%	Neutral
R98T	60%	Neutral	T223N	83%	Neutral	W348R	76%	Deleterious	A497S	83%	Neutral
E99A	65%	Neutral	Y224C	76%	Deleterious	K349E	60%	Neutral	A497T	64%	Deleterious
E99D	83%	Neutral	A225S	83%	Neutral	K349N	83%	Neutral	A497V	72%	Deleterious
E99G	65%	Neutral	A225T	83%	Neutral	K349Q	75%	Neutral	Y498C	87%	Deleterious
A100S	74%	Neutral	A225V	60%	Neutral	K349R	83%	Neutral	Y498H	87%	Deleterious
V101F	76%	Deleterious	W227L	63%	Neutral	K349T	60%	Neutral	M500K	55%	Deleterious
V101I	83%	Neutral	W227R	76%	Deleterious	F350L	60%	Neutral	M500L	83%	Neutral
T103S	87%	Deleterious	H228R	83%	Neutral	F350S	76%	Deleterious	M500R	61%	Deleterious
L107S	87%	Deleterious	H228Y	63%	Neutral	A353S	74%	Neutral	M500T	83%	Neutral
F111L	87%	Deleterious	S230P	72%	Deleterious	A353T	75%	Neutral	M500V	83%	Neutral
S112A	87%	Deleterious	S230T	74%	Neutral	A353V	63%	Neutral	M501I	75%	Neutral
T113A	76%	Deleterious	I231M	74%	Neutral	Q354H	83%	Neutral	M501L	83%	Neutral
T113I	65%	Deleterious	I231T	83%	Neutral	Q354R	74%	Neutral	M501T	60%	Neutral
G114C	87%	Deleterious	I231V	83%	Neutral	P355L	87%	Deleterious	M501V	74%	Neutral
G114S	87%	Deleterious	G232R	87%	Deleterious	P355S	87%	Deleterious	I502L	74%	Neutral
G114V	87%	Deleterious	F233L	75%	Neutral	C356Y	55%	Deleterious	I502T	83%	Neutral
V115A	74%	Neutral	D234A	87%	Deleterious	S357G	83%	Neutral	I502V	83%	Neutral
V115F	61%	Deleterious	D234G	87%	Deleterious	S357I	75%	Neutral	S503A	74%	Neutral
V115I	83%	Neutral	D234N	87%	Deleterious	S357N	83%	Neutral	S503L	65%	Neutral
N116D	83%	Neutral	D234Y	87%	Deleterious	D358E	83%	Neutral	S503P	87%	Deleterious
L117I	63%	Neutral	Y235C	87%	Deleterious	D358G	83%	Neutral	S503T	83%	Neutral
L117P	76%	Deleterious	V236F	87%	Deleterious	D358N	83%	Neutral	A504D	87%	Deleterious
V118F	87%	Deleterious	V236I	74%	Neutral	D358Y	55%	Deleterious	A504S	52%	Deleterious
V118I	65%	Neutral	V236L	83%	Neutral	K359R	83%	Neutral	A504T	87%	Deleterious
A119S	51%	Deleterious	N238D	87%	Deleterious	A360S	60%	Neutral	A504V	61%	Deleterious
A119T	51%	Deleterious	N238S	87%	Deleterious	A360T	51%	Deleterious	A504Y	87%	Deleterious
A119V	83%	Neutral	P239L	87%	Deleterious	A360V	75%	Neutral	F506S	87%	Deleterious
V120A	74%	Neutral	P239S	76%	Deleterious	Y361C	52%	Deleterious	S507G	60%	Neutral
V120I	74%	Neutral	F240L	83%	Neutral	Y361F	71%	Neutral	S507I	76%	Deleterious
V120L	74%	Neutral	F240S	72%	Deleterious	I363T	51%	Deleterious	L508F	83%	Neutral
P121L	72%	Deleterious	M241I	83%	Neutral	I363V	83%	Neutral	W509R	87%	Deleterious
P121S	65%	Deleterious	M241L	83%	Neutral	E364A	61%	Deleterious	V510A	61%	Deleterious
T122A	65%	Neutral	M241T	63%	Neutral	E364D	75%	Neutral	V510F	87%	Deleterious
Y124C	83%	Neutral	M241V	83%	Neutral	E364K	83%	Neutral	Y511H	72%	Deleterious
Y124F	83%	Neutral	I242L	60%	Neutral	E365G	83%	Neutral	K512R	83%	Neutral
V125F	83%	Neutral	I242S	87%	Deleterious	E365K	83%	Neutral	Q513R	60%	Neutral
V125I	68%	Neutral	I242T	79%	Deleterious	L366F	60%	Neutral	F514L	72%	Deleterious
P128H	75%	Neutral	I242V	83%	Neutral	F367C	63%	Neutral	D515A	87%	Deleterious
P128S	74%	Neutral	V244I	83%	Neutral	F367L	83%	Neutral	D515G	87%	Deleterious
N129A	74%	Neutral	Q245L	87%	Deleterious	F367S	83%	Neutral	T516A	83%	Neutral
N129D	83%	Neutral	W247C	87%	Deleterious	S369F	68%	Neutral	T516I	83%	Neutral
N129G	83%	Neutral	G248C	87%	Deleterious	S369Y	74%	Neutral	T516N	72%	Deleterious
N129S	83%	Neutral	G248D	72%	Deleterious	Y370C	87%	Deleterious	N521S	83%	Neutral
N129Y	83%	Neutral	G248S	65%	Deleterious	Y370F	63%	Neutral	T522A	61%	Deleterious
N130S	65%	Neutral	G248V	87%	Deleterious	A371S	83%	Neutral	T522I	60%	Neutral
T131I	64%	Deleterious	F249L	51%	Deleterious	A371T	83%	Neutral	T524I	72%	Deleterious
D132A	75%	Neutral	F249S	72%	Deleterious	A371V	83%	Neutral	R525I	63%	Neutral
D132E	74%	Neutral	T250A	52%	Deleterious	T372I	83%	Neutral	R525K	83%	Neutral
D132G	83%	Neutral	T250I	74%	Neutral	H373Q	72%	Deleterious			
D132H	74%	Neutral	G251S	87%	Deleterious	H373Y	52%	Deleterious			
D132N	83%	Neutral	N252K	74%	Neutral	S374A	83%	Neutral			
D132Y	83%	Neutral	N252S	83%	Neutral	S374F	63%	Neutral			
F133L	83%	Neutral	Q254H	72%	Deleterious	S374Y	63%	Neutral			
S134A	75%	Neutral	Q254K	61%	Deleterious	D375A	61%	Deleterious			
S134F	63%	Neutral	Q254L	60%	Neutral	D375E	83%	Neutral			
S134Y	60%	Neutral	S255C	63%	Neutral	D375G	76%	Deleterious			
R135G	74%	Neutral	S255G	74%	Neutral	D375H	63%	Neutral			
R135I	65%	Neutral	S255I	83%	Neutral	D375N	74%	Neutral			
R135K	83%	Neutral	S255N	60%	Neutral	D375Y	87%	Deleterious			
V136F	76%	Deleterious	S255R	61%	Deleterious	K376E	63%	Neutral			
V136I	83%	Neutral	D258A	76%	Deleterious	K376N	83%	Neutral			
S137G	83%	Neutral	D258G	87%	Deleterious	K376R	83%	Neutral			
S137I	83%	Neutral	D258N	52%	Deleterious	F377L	87%	Deleterious			
A138S	65%	Neutral	D258Y	87%	Deleterious	F377Y	87%	Deleterious			
A138V	76%	Deleterious	L259M	83%	Neutral	T378A	83%	Neutral			
K139R	83%	Neutral	Y260C	60%	Neutral	T378I	83%	Neutral			

Table 14: List of co-mutated sequences (CS) informatics. Nine bold marked CS were co-mutated sequences due to deleterious (DD type) mutations only.

Co-mutated sequences (CS) due to co-mutations		
CS1(P43L, Q343L)	**CS41**(V263F, P412H)	CS81(N129D, P355S)
CS2(P43L, V510A)	**CS42**(P412H, M169I)	CS82(N129D, S374F)
CS3(P43L, G251S)	CS43(P412H, K147R)	CS83(N129D, L383I)
CS4(P43L, H486L)	CS44(P412H, K311N)	CS84(N129D, E204D)
CS5(P43L, V328F)	**CS45**(M169I, A119S)	CS85(N129D, T219I)
CS6(P43L, T250I)	CS46(P140L, A371V)	CS86(N129D, N67T)
CS7(P43L, I74T)	CS47(P140L, A344S)	CS87(N129D, P443S)
CS8(P43L, D144Y)	CS48(A344S, L493F)	CS88(N129D, K304R)
CS9(P43L, A138V)	CS49(L493F, S374A)	CS89(N129D, P239S)
CS10(P43L, P140S)	CS50 (H26Y, P24L)	CS90(P451S, K304T)
CS11(P43L, A100S)	CS51(H26Y, R525K)	CS91(P451S, P443S)
CS12(P43L, L177F)	CS52(H26Y, A274S)	CS92(P451S, S221A)
CS13(P43L, E453D)	CS53(H26Y, V136F)	CS93(P451S, K469R)
CS14(P43L, L157F)	CS54(P46S, G248V)	CS94(P451S, V263F)
CS15(P43L, M501I)	**CS55**(T113I, V290F)	CS95(P451S, I45T)
CS16(P140S, D432G)	CS56(R163C, S461L)	CS96(M58I, T16I)
CS17(L177F, G6R)	CS57(A119V, V236F)	CS97(M58I, Y260C)
CS18(E453D, E347G)	CS58(R213L, V287F)	**CS98(Q22H, C208S, T524I)**
CS19(L157F, H373Y)	**CS59**(M153I, Q254K)	**CS99(Q22H, C208S, S481I)**
CS20(M501I, T31I)	CS60(N129D, K13R)	CS100(Q22H, E36G)
CS21(M501I, A220V)	CS61(N129D, D48N)	CS101(Q22H, V381I)
CS22(M501I, V125F)	CS62(N129D, H486R)	CS102(Q22H, C208S)
CS23 (M501I, K349N)	CS63(N129D, G114C)	CS103(A394V, V182L)
CS24(K349N, A225S)	CS64(N129D, T131I)	CS104(A394V, V120A)
CS25(K349N, V421I)	CS65(N129D, D324Y)	CS105(A394V, S369F)
CS26(V421I, N238S)	CS66(N129D, S255I)	CS106(A394V, C208S)
CS27(V125F, I166V)	CS67(N129D, R212K)	CS107(P142S, A353V)
CS28(T31I, V341A)	CS68(N129D, A360V)	**CS108**(P142S, D222Y)
CS29(T31I, M315I)	CS69(N129D, V4L)	CS109(M62L, N63K)
CS30(M315I, M62I)	CS70(N129D, V182I)	**CS110**(F217V, Y511H)
CS31(M315I, A319S)	CS71(N129D, A360S)	CS111(D324E, L409F)
CS32(V125F, V381L)	CS72(N129D, P158S)	CS112(M72I, V341I)
CS33(V381L, T97I)	CS73(N129D, L366F)	CS113(M153V, K155R)
CS34(V381L, H373Y)	CS74(N129D, R476S)	CS114(S221P, A471V)
CS35(V381L, V14L)	CS75(N129D, D145G)	CS115(L157F, A225V)
CS36(V381L, V263F)	CS76(N129D, D258N)	CS116(T219I, E204D)
CS37(V381L, D496Y)	CS77(N129D, M49V)	CS117(T524I, I15T)
CS38(V381L, H455Y)	CS78(N129D, T524I)	CS118(M49V, T524I)
CS39(H455Y, N176D)	CS79(N129D, T161)	CS119(R212K, A323S)
CS40(D496Y, A320V)	CS80(N129D, S450G)	CS120(N129D, T16I)