Non-synonymous Mutations of SARS-Cov-2 Leads Epitope Loss and Segregates its Varaints

Aayatti Mallick Gupta
S. N. Bose National Centre for Basic Sciences
Sukhendu Mandal (sukhendu1@hotmail.com)
University of Calcutta https://orcid.org/0000-0002-7752-0982

Research Article

Keywords: SARS-Cov-2, COVID-19, non-synonymous mutation, epitope loss, phylogenomics

Posted Date: May 21st, 2020

DOI: https://doi.org/10.21203/rs.3.rs-29581/v1

Abstract

The non-synonymous mutations of SARS-Cov-2 isolated from across the world have been identified during the last few months. The surface glycoprotein spike of SARS-Cov-2 forms the most important hotspot for amino acid alterations followed by the ORF1a/ORF1ab poly-proteins. It is evident that the D614G mutation in spike glycoprotein and P4715L in RdRp is the important determinant of SARS-Cov-2 evolution since its emergence. P4715L in RdRp, G251V in ORF3a and S1498F of Nsp3 is associated with the epitope loss that may influence pathogenesis caused by antibody escape variants. Phylogenomics distinguished the ancestral viral samples from China and most part of Asia, isolated between Dec, 2019 to Feb, 2020 and the evolved variants isolated from Europe and Americas from Mar, 2020 to April, 2020. The evolved variants have been found to predominant globally with the loss of epitopes from its proteins. These have implications for SARS-Cov-2 transmission, pathogenesis and immune interventions.

Background

The current outbreak of COVID-19 caused by SARS-Cov-2 enforced the greatest global health and a socio-economic threat to mankind [1]. It was first reported in late December 2019 from Wuhan, China [2-4], become an epidemic and rapidly spread across the globe to become a pandemic with devastating impact [5]. At the end of January 2020 India reported its first case of COVID-19 from the state of Kerala. Two months since its outbreak by the middle of March 2020, Europe and North America have become the new epicenters to the pandemic with remarkable expansion of this disease with a huge number of fatalities. A total of 3435894 positive cases and 239604 deaths worldwide and 42533 confirmed incidence and 1373 deaths from India has been reported by the WHO situation report, 4th May 2020.

SARS-Cov-2 is a novel member of the beta-coronavirus genus with single-stranded positive-sense RNA. They have similarities with the severe acute respiratory syndrome coronavirus (SARS-CoV) and to several bat coronaviruses [6]. SARS-Cov-2 comprises of around 29903 nucleotides organized into specific genes characteristics within its genome. In the 5' region more than two-thirds of the genome comprises a set of non-structural proteins (Nsp) produced as cleavage products of the ORF1a and ORF1ab viral polyproteins [7] assemble to facilitate viral replication and transcription. RNA-dependent RNA polymerase (RdRp, also known as Nsp12), is the key component that regulates the synthesis of viral RNA with the assistance of Nsp7 and Nsp8 as co-factors [8]. The 3' region consists of genes encoding structural proteins including surface (S), envelope (E), membrane (M), and nucleocapsid (N) proteins. Surface glycoprotein spike is involved in the interaction with the host's receptor, Ace2, further ingress, and forms one of the vital factors for rapid human-to-human transmission [9]. Additionally, 6 accessory proteins are encoded by ORF3a, ORF6, ORF7a, ORF7b, and ORF8 genes (Fig. 1a). Compared to SARS-CoV emerged in 2002 and MERS-CoV in 2012, SARS-CoV-2 exhibits faster human transmission [10] and leads to the declaration of a worldwide public health emergency by WHO [2,3]. Three factors that make SARS-Cov-2 associated disease more infectious include higher transmissibility, high mortality, and humans have no prior immunological history against it.

In general RNA viruses are vulnerable to a high rate of mutations [11] which may be correlated with the geographical region-specific virulence of virus variants. The rapid global spread of SARS-Cov-2 provides an ample opportunity for natural selection to act upon rare but favorable mutations. There might have complex interplay between amino acids that can confer immune resistance to the virus and the fitness landscape of the particular variant. Mutation of single amino acid within an antigenic determinant or epitope can potentially overcome the antibody recognition. The ability to identify epitope is vital to combat the infection through antibody-mediated immunity and also essential in several biomedical applications like rational vaccine design, disease diagnostic, and immune-therapeutics [12,13]. Non-synonymous mutation leading to the loss of epitope allows escaping antibody immunity. Such antibody escape could also be mediating individuals susceptible to secondary infection. The nature and locations of epitope losses from the various proteins of SARS-Cov-2 due to non-synonymous mutations are the main focus of the present study. The immediate and continuous release of complete genome sequences of SARS-Cov-2 from samples of diverse geographical regions helped scientists to monitor the rapid evolution of the SARS CoV-2 to gain insights into the pattern and dynamics global spread of COVID-19. We have performed a comparative genome analysis of the viral samples from a varied geographical location with special reference to India. The mutations in the SARS-Cov-2 genome that cause loss of epitopes are following a specific pattern of its emergence. Based on such non-synonymous mutations in the SARS-Cov-2 genome, the initial viral isolates are under different clusters than the later evolved one. For example, the Indian isolates of SARS-Cov-2 clearly distributed and depicted that the initial samples from Kerala are having similarity with the Wuhan isolates and the rest are predominantly found within isolates originated from patients having travel history from Italy. These findings have important implications to understand the impact of new emerging mutations in the pathogenesis and immune evasion of SARS-CoV-2.

1. Materials And Methods

1.1. Genome analysis

A total of 87 SARS-Cov-2 high-quality complete genome sequences sequenced across the world and reported from January to April 2020 were accessed from GISAID Epicov [14] platform (Supplementary Table1). The infection load and geographical locations from various continents viz. Asia, Africa, Europe, North, Central, and South Americas, and Oceania were considered for complete genome selections. Each of the representative genome sequences from different COVID 19 affected countries of these continents and eleven genome sequences reported from China was used for this study. All the Indian isolates available till the middle of April 2020 were taken for genome analysis to unravel the nature of SARS-Cov-2. All sequences were uniformly annotated using the RAST server [15]. Protein domain analysis was conducted with Pfam [16].

1.2. Multiple sequence alignment

Clustal omega [17] has been employed to align ORF1a, ORF1ab, spike, ORF3a, E, M, ORF6, ORF7a, ORF7b, ORF8 and N proteins of SARS-Cov-2 sequences originated from all affected countries. Thus the non-synonymous mutations were identified from the multiple sequence alignment.

1.3. Effect of amino acid alterations

PROVEAN algorithm was used to predict the functional effect of non-synonymous mutations [18]. Change in epitope due to amino acid substitutions has been predicted from the Immune epitope database analysis (IEDB-AR) with 0.5 threshold and default settings [19]. Change of secondary structure of the respective proteins due to such amino acid alterations have been done with PSIPRED [20]. Prediction on protein structural stability changes upon mutation using Normal Mode Analysis has been done with the Dynamut server [21]. Visual analysis of the structure has been done with PYMOL molecular graphics system.

1.4. Phylogenomic analysis

Phylogenetic analysis of whole-genome sequences was inferred with REALPHY 1.12 [22]. It uses the maximum likelihood algorithm with the GTR model to build the tree. The resulted tree was visualized and annotated using FigTree (http://tree.bio.ed.ac.uk/software/figtree/).

2. Results

2.1. Spike glycoprotein is the most important hotspot of amino acid substitution in SARS-Cov-2

The study showed 47 non-synonymous mutations from the genome sequence of 87 independent SARS-Cov-2 isolated from more than fifty representative countries of five continents. The total number of neutral mutations was almost equal to that of the deleterious mutations (Table 1). Spike glycoprotein was found to be the most vital hotspot of amino acid alterations among proteins of SARS-Cov-2. The surface protein spike mediates receptor recognition through its RBD (receptor binding domain) region with human Ace2. 13 different mutations of spike protein were observed of which D614G was most notable and exist among 37 samples, especially among isolates from the European countries. However, the total number of neutral mutations in spike glycoprotein was higher than the deleterious mutations. However, there was no amino acid alteration observed in the RBD region of the spike. This might be just because all the samples originated from the patient, which means the virus went through successful pathogenesis following the interaction to the host receptor through the spike. Thus there is a high chance that the virus might undergo mutation in RBD of spike protein and become non-infective or negligible in number to be highlighted for such mutation mapping. After spike glycoprotein, it is the orf1a polyprotein that forms the major site of amino acid alterations. Orf1a expresses the viral replicase and protease enzyme. The dominance of deleterious mutation was found high in this region. S1498F was the most frequent mutation at the multi-domain essential replication/transcription complex- Nsp3 of ORF1a and reflected among multiple Indian isolates with travel history to Italy. Nucleocapsid N protein conferred 7 non-synonymous mutations of which 4 are neutral and 3 are deleterious. An incident of the concurrence

of R203K and G204R was found in Belgium, Brazil, Peru, Mexico, Nigeria, Switzerland, and Vietnam. Similar mutations also have been spotted in 3 samples of Indian isolates having a travel connection to Italy. ORF1ab consisted of an equal number of neutral and deleterious mutation effects. P4715L variant of ORF1ab was remarkable to occur among 37 samples in the RNA dependent RNA polymerase (RdRp) site. This prime spot for mutation is especially seen among the European samples (Supplementary table 1). The Indian isolates from patients with European travel history elicited such kind of variation deviating from the ancestral samples from China. ORF3a showed a more deleterious effect of amino acid substitutions. The incidence of occurrence of G251V was high among various samples irrespective of geographical locations. However, Indian isolates did not elicit any mutation at ORF3a. The change of amino acid L84S in ORF8 was observed among the viral samples of more than 12 countries. Such a mutation was obtained from affected countries of varied geographical locations. Primary samples isolated in India from Kerala in January 2020 bear such mutation. V74F is the deleterious mutation found in ORF7b among an isolate from Kuwait and five Indian isolates sampled at Iran in March 2020. It can be called that V74F variation in ORF7b is primarily restricted to the Asian continent. Membrane glycoprotein consists of a neutral and deleterious mutation not having any amino acid alteration on Indian samples. Amino acid alterations in envelope protein and two other accessory proteins ORF6 and ORF7a were not found among the isolates chosen for the present study. The relative occurrence of non-synonymous mutations among gene products of SARS-Cov-2 has been shown in Fig. 1b.

2.2 Loss of epitope for non-synonymous mutation

At least there are three regions where amino acid substitutions in the protein sequence led to the loss of epitope. P4715L substitution in RdRp protein from the ORF1ab region is linked with FPPTSFG epitope loss from the site (Fig. 2a). Wild type RdRp consists of seven epitope regions that are reduced to six due to the amino acid substitutions. P to L mutation is neutral which has been found to elicit the change from a ‘turn’ to ‘helix’ in RdRp secondary structure, adjacent to the site of the epitope (Supplementary Fig. S1). The second epitope loss is associated with G251V in ORF3a (Fig. 2b). Six putative epitopes in the wild-type ORF3a is replaced by five in the mutant variants. The third loss of epitope S1498F is attached to non-structural protein Nsp3 in ORF1a among Indian isolates. In this case, it is observed that the epitope loss is partial. The substitution of S by F in the region causes the elimination of five residues ‘YKDWS’ from the wild type variant (Fig. 2c). Such loss of epitope might allow the new variant to escape interaction with host antibodies and influence disease profile by evading the antibody-mediated neutralization.

2.3. Structural impact due to P4715L in RdRp and D614G in Spike protein

To study the effect of the mutation on the tertiary structure of RdRp and spike protein PDB ID: 6M71, chain A, PDB ID: 6VSB, chain A have been used as wild type structures, respectively. P4715L in RdRp accounts for stable conformation change confirmed by binding conformational enthalpy change (ΔΔG = 1.540 kcal/mol). The vibrational entropy changes (ΔΔSVib ENCoM = -4.074 $kcal.mol^{-1}.K^{-1}$) for this mutation signify decrease in molecular flexibility (Fig S1a and 5b). The shape of the epitope due to Pro to

Leu substitution becomes distorted ellipse-like as elicited from the visual analysis of the tertiary structure change due to mutation of RdRp (supplementary Fig. S2). In spike glycoprotein, D614G is favored by the attainment of molecular stability (ΔΔG = 1.128 kcal/mol) and vibrational entropy changes (ΔΔSVib ENCoM = -4.531 kcal.mol^{-1}.K^{-1}) leading to decrease in molecular stability (supplementary Fig S1c and 5d).

2.4. Coexistence of P4715L in RdRp and D614G in Spike protein among European and American samples

It is interesting to found that L4715 RdRp and D614G spike variants show co-occurrence (Fig. 3). These variants from two completely different proteins are highly correlated. The same viral sample is carrying both the mutation at any time. Such modifications in spike and RdRp are adopted more in SARS-Cov-2 isolated from patients of Europe and Americas than those isolated from Asia. However, the initial isolates from China lack such variation completely. The pandemic burden of COVID-19 in India consists of both types of SARS-Cov-2. Indian patients with travel history from Europe and Americas are mainly affected by such variants. In Africa only two samples are considered and both of them showed such a new variant of spike and RdRp. Thus it can be anticipated that these two variants play a major role in outrageous infection associated with various parts of Europe and Americas.

2.5. The more evolved form of SARS-Cov-2 is indicated from phylogenomic analysis

The adaptive evolution of this novel pathogen is not biased with geographic location but is related to the non-synonymous mutation located in spike glycoprotein and RdRp of ORF1ab. P4715L RdRp and D614G spike is the determining factor to cluster the different SARS-Cov-2 isolates into different clades. The phylogenetic tree using whole-genome sequence elicits two types of clades- the green represented by the ancestral type from China and red with the evolved variant in spike and RdRp (Fig. 4). Thus the red group is the mutant form showing the loss of epitope among various infected countries. The green group is more prominent in China and other parts of Asia and the red group mainly belongs to Europe and America. From the Indian perspective, it is very clear that Indian samples isolated from patients with the contact or travel history linked to Wuhan are green variants and the rest linked with other than Asia, especially Europe and America, are the red variants. The red variants seem to be more infectious than the green and the mutations with epitope loss may have a definite role in this

3. Discussion

The impact of epitope loss due to non-synonymous mutation is the biggest and unique concern highlighted in this study. It could potentially be linked to immune evasion and thus higher viral spread and pathogenesis. The real-time evolution of SARS-Cov-2 has been tracked from the available whole genome sequences around the globe through the phylogenomic analysis. It has been found that there are many point mutations incorporated in the genome of SARS-Cov-2 from the variants associated with the first outbreak in China to the present variants isolated in Europe. It is further interesting to note that all the

samples isolated after the middle of March 2020 to 30th April 2020 show the major prevalence of epitope loss due to such amino acid alterations. Mutations that appeared at multiple times include D614G in spike glycoprotein and P4715L in RdRp. Such mutations occurred primarily as the virus started perpetuating outside China during the second sharp infection and morbidity in Italy, Spain, and other parts of Europe. The present analysis clearly justifies the concurrence of these two mutation sites in the same viral sample. Increased transmission of such mutations might have a selective advantage for positive selection Conformation changes in SARS-Cov-2 protein structure due to much non-synonymous mutation and epitope loss could be immensely interesting both for plasma mediated therapy or serological detection of COVID-19. However, it can be anticipated that the replacement of negatively charged Asp into non-polar Gly in D614G mutation of spike protein consequences from the replacement of hydrophilic to the hydrophobic residue. Furthermore the flexibility has also been compromised due to such mutation as reflected from the vibrational entropy changes. Thus the loss of acidic residue and impairment in structural flexibility might play a role in altered recognition towards human Ace2 receptor recognition and fusion of viral membrane. Structural analysis has indicated that the D to G mutation at 614th position in spike protein does not cause significant variation as both of them attain sheet-like conformation in their secondary structure. In the tertiary structure it appeared that such mutation stabilized the structure better however decreased the structural flexibility as accounted from binding conformational enthalpy change. The phenomenon of epitope loss due to P4715L in ORF1ab polyprotein is supported because Leu (L) is the small hydrophobic amino acid that shows less preference in the epitope residues [23-26]. The change of secondary structure from loop to helix linked to P4715L supports epitope loss as epitope antibody residues are enriched by loops and depleted of strands and helixes [23,24]. This is logical due to the substitution of helix breaker –Pro (P) with a most stabilizing residue in alpha helix, Leu (L) might cause structural stabilization. The ellipse-like shape of the epitope [24,26] 'FPPTSFG' becomes distorted due to amino acid alteration leading to the loss of epitope. Thus the change of shape of the epitope also supports the cause behind the loss of antigenic determinant. Structural changes in RdRp might influence viral replication as it is one of the key players in viral replications. It can be hypothesized, that the decrease in flexibility of the overall structure of spike and RdRp due to mutations and epitope loss made them less accessible to antibody causing less responsive immunity by the host. Such stabilizing mutation might thus account for positive selection and adaptive advantage as when introduced to new variant sit rapidly becomes the dominant form. The epitope loss linked with G251V in ORF3a might cause a change in flexibility due to the substitution of small hydrogen moiety of Gly (G) to the bulky branched side chain in Val (V). The tendency for epitopes to be depleted of small hydrophobic amino acids like Val [23-26] supports the epitope loss associated with G251V. Such changes in the ORF3a-viroporin domain require further experimental outcomes to precisely understand the effect exerted.

While trying to unravel the role of non-synonymous mutation from the Indian perspective, it has been found that the Indian SARS-Cov-2 samples isolated from the patient of Kerala state with travel history from Wuhan on January lack D614G mutation in their spike glycoprotein but additionally shows a neutral mutation R416I and an alteration of A930V with deleterious effect. Deletion is also found at Y145 in the

initial viral sample hCoV-19/India/1-27/2020 during the report of onset of infection in India. The D to G mutation at 614th position among Indian samples has been found with European travel connections. Increased infectivity might be consistent with rapid spread, and also the association of higher viral load with G614 that we observed in more infected countries. For P4715L mutation in RdRp, a mixed outcome has been noticed. P4715 with the epitope 'FPPTSFG' remains conserved to all the samples isolated from patients without travel history outside Asia. L4715 with the loss of epitope is found in the samples with travel history from Italy. Indian isolates from European origin show a new mutation S1498F in the papain-like protease domain of Nsp3 in the ORF1a region. It is responsible for releasing Nsp1, Nsp2, and Nsp3 from the N-terminal region of polyproteins 1a/1ab [27]. Ser (S) to Phe (F) mutation led to a shift from polar hydrophilic to the non-polar hydrophobic group. The residues in epitope are enriched by polar amino acids and depleted of hydrophobic amino acid compared to residues of non-epitope n residues [23-26,28,29] supports the observation of epitope loss from this site. As the properties of these two amino acids are different, the mutations might presumably modify the replication/transcription function of Nsp3. V74F mutation appears in the ORF7a accessory protein among isolates from the COVID-19 patients of India who migrated from Iran. This accessory protein is suggestive to have a potential role in binding with the IL-1 receptor of the host cell [30]. Such a mutation from V to F is deleterious and may have an impact as the side chain isopropyl group is substituted by a phenyl group. R203K and G204R in nucleocapsid protein are found in SARS-Cov-2 of three Indian patients who traveled from Italy in April 2020. Thus the 32 Indian samples studied so far consist of mutation at 10 sites including a deletion at spike protein. Membrane glycoprotein and ORF3a is uniform among all the Indian isolates. Multiple emergent variants of viral genomes both from Europe and China are circulating in the Indian population. Sequences submitted from Kerala are similar to the original Wuhan virus while the rest are similar to Italy. There can be a better understanding through large-scale global sequencing studies which clears the most perpetuating SARS-Cov-2 in the Indian population. Meanwhile, understanding how mutations could lead to loss of epitope and how this in turn related to probable immune evasion would be an important topic of research to combat the pandemic.

Declarations

Acknowledgments

We are thankful to the authors, generating and submitting laboratories of the sequences from GISAID's EpiCoV™ database. AMG thanks SNBNCBS for fellowship. The authors highly acknowledge Prof Jaydeb Chakrabarti for proofreading the manuscript. SM supervised the work. AMG has analyzed and interpreted the data. Both the authors are involved in the drafting of the manuscript.

Disclosure statement

The authors wish to declare that they do not have any conflict of interest.

References

[1] Poon LLM, Peiris M. Emergence of a novel human coronavirus threatening human health. Nat Med. 2020;26:317–319.

[2] Chan JF, Yuan S, Kok KH, To KKW, Chu H, Yang J, et al. A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to person transmission: A study of a family cluster. Lancet. 2020; 395:514–523.

[3] Chen N, Zhou M, Dong X, Qu J, Gong F, Han Y, et al. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: A descriptive study. Lancet. 2020;395: 507–513.

[4] Wu F, Zhao S, Yu B, Chen YM, Wang W, Song ZG, et al. A new coronavirus associated with human respiratory disease in China. Nature. 2020;579:265–269.

[5] Cucinotta D, Vanelli M. WHO declares COVID-19 a pandemic. Acta Bio-Medica Atenei Parm. 2020;91:157-160.

[6] Zhou P, Yang XL, Wang XG, Hu B, Zhang L, Zhang W, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature. 2020;579:270–273.

[7] Ziebuhr J. The coronavirus replicase. Curr. Top. Microbiol. Immunol. 2005;287:57–94.

[8] Subissi L, Posthuma CC., Collet A, Zevenhoven-Dobbe JC, Gorbalenya AE, Decroly E, et al. One severe acute respiratory syndrome coronavirus protein complex integrates processive RNA polymerase and exonuclease activities. Proc. Natl. Acad. Sci. U.S.A. 2014;111:E3900–E3909.

[9] Li W, Moore MJ, Vasilieva N, Sui J, Wong SK, Berne MA, et al. Angiotensin converting enzyme 2 is a functional receptor for the SARS coronavirus. Nature. 2003; 426:450–454.

[10] Gralinski LE, Menachery VD. Return of the Coronavirus: 2019-nCoV. 2. Viruses. 2020;12:135.

[11] Tang X, Wu C, Li X, Song Y, Yao X, Wu X, et al. On the origin and continuing evolution of SARS-CoV-2. Natl Sci Rev. 2020;nwaa036

[12] Gershoni JM, Roitburd-Berman A, Siman-Tov DD, Tarnovitski Freund N, Weiss Y. Epitope mapping: the first step in developing epitope-based vaccines. BioDrugs. 2007;21:145–156.

[13] Irving MB, Pan O, Scott JK. Random-peptide libraries and antigen-fragment libraries for epitope mapping and the development of vaccines and diagnostics. Curr Opi Chem Biol. 2001;5:314–324.

[14] Shu Y, McCauley, J. GISAID: Global initiative on sharing all influenza data - from vision to reality. Euro Surveill. 2017;22:30494.

[15] Brettin T, Davis JJ, Disz T, Edwards RA, Gerdes S, Olsen GJ, et al. *RASTtk: A modular and extensible implementation of the RAST algorithm for building custom annotation pipelines and*

annotating batches of genomes. Sci Rep. 2015;5:8365.

[16] Gebali-el S, Mistry J, Bateman A, Eddy SR, Luciani A, Potter SC, et al *The Pfam protein families database in 2019* NAR. 2019;47:D427-432.

[17] Sievers F, Wilm A, Dineen D, Gibson TJ, Karplus K, Li W, et al. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol Syst Biol. 2011;7:539.

[18] Choi Y, Chan AP. PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. Bioinfo. 2015;31:2745–2747.

[19] Zhang Q, Wang P, Kim Y, Haste-Andersen P, Beaver J, Bourne PE, et al. Immune epitope database analysis resource (IEDB-AR). NAR 2008;36:W513-518.

[20] Jones DT. Protein secondary structure prediction based on position-specific scoring matrices. J. Mol. Biol. 1999;292:195-202.

[21] Rodrigues CH, Pires DE, Ascher DB DynaMut: predicting the impact of mutations on protein conformation, flexibility and stability. NAR. 2018;46(W1):W350–W355.

[22] Guindon S, Dufayard JF, Lefort V, Anisimova M, Hordijk W, Gascuel O. "New algorithms and methods to estimate maximum-likelihood phylogenies: assessing the performance of PhyML 3.0." Systematic Biology. 2010;59:307-321.

[23] Ofran, Y., Schlessinger A, Rost B. Automated identification of complementarity determining regions (CDRs) reveals peculiar characteristics of CDRs and B cell epitopes. J Immun 2008;181:6230–6235.

[24] Rubinstein ND, Mayrose I, Halperin D, Yekutieli D, Gershoni JM, Pupko T. Computational characterization of B-cell epitopes. Mol Immuno. 2008;45, 3477–3489.

[25] Sun J, Xu T, Wang S, Li G, Wu D, Cao Z. Does difference exist between epitope and non-epitope residues? Analysis of the physicochemical and structural properties on conformational epitopes from B-cell protein antigens. Immuno Res. 2011;7:1–11.

[26] Kringelum JV, Nielsen M, Padkjær SB, Lund O. Structural analysis of B-cell epitopes in antibody:protein complexes. Mol Immunol. 2013;53:24-34.

[27] Harcourt BH, Jukneliene D, Kanjanahaluethai A, Bechill J, Severson KM, Smith CM, et al. Identification of severe acute respiratory syndrome coronavirus replicase products and characterization of papain-like protease activity. J Virol. 2004;78: 13600-12.

[28] Andersen PH, Nielsen M, Lund O. Prediction of residues in discontinuous B-cell epitopes using protein 3D structures. Prot Sci. 2006;15:2558–2567.

[29] Zhao L, Li J. Mining for the antibody-antigen interacting associations that predict the B cell epitopes. BMC Struc Bio. 2010;10:S6.

[30] Hänel K, Stangler T, Stoldt M, Willbold D. Solution structure of the X4 protein coded by the SARS related coronavirus reveals an immunoglobulin like fold and suggests a binding activity to integrin I domains. J Biomed Sci. 2006;13:281-93.

Table

Table1: The non-synonymous substitutions and their effects among various global isolates of SARS-Cov-2 included in this study. Indian isolates are indicated in red.

	Variant	Infected country	Mutation effect
ORF1a polyprotein	D58E	New Zealand	Deleterious
	L952P	China(hCoV-19/Shandong/LY003/2020)	Neutral
	E955K	China(hCoV-19/Tianmen/HBCDC-HB-07/2020)	Deleterious
	S1498F	Multiple Indian isolates	Deleterious
	N1559T	Russia	Neutral
	A3203V	USA	Neutral
	G4227R	Russia	Deleterious
	A4297G	Mexico	Deleterious
	F4304L	Sweden	Deleterious
ORF1ab polyprotein	P4715L	Multiple countries isolates[$]	Neutral
	Y232C	Australia,USA	Deleterious
	F1657L	New Zealand	Neutral
	A1906V	Canada	Deleterious
	V1973L	New Zealand	Neutral
	G2374R	France	Deleterious
Spike protein	Y145del	India(hCoV-19/India/1-27/2020)	Neutral
	N354D	China(hCoV-19/Shenzhen/SZTH-004/2020)	Neutral
	D364Y	China(hCoV-19/Shenzhen/SZTH-004/2020)	Deleterious
	R416I	India(hCoV-19/India/1-27/2020)	Neutral
	S438F	India(hCoV-19/India/763/2020,777/2020)	Neutral
	Y508H	France	Neutral
	D614G	Multiple countries isolates[$]	Neutral
	Q675H	Scotland	Neutral
	T791I	Taiwan	Neutral
	F797C	Sweden	Deleterious
	A930V	India(hCoV-19/India/1-31/2020)	Deleterious
	I1216T	China(hCoV-19/Shanghai/SH0067/2020)	Deleterious
	P1263L	England	Neutral
ORF3a	A31T	Hungary	Neutral
	Q57H	Russia, Congo,Saudi Arabia	Deleterious
	V88L	Cambodia	Neutral
	H93Y	Wales	Deleterious
	G196V	Chile, Georgia	Deleterious
	G251V	France, Sweden, Australia, China(hCoV-19/HongKong/CUHK1/2020)	Deleterious
Membrane glycoprotein	D3G	Finland	Neutral
	T175M	Belgium, Brazil	Deleterious
ORF7a	V74F	Kuwait,India(hCoV-19/India/1073/2020,1093/2020,1100,2020,1115/2020,1063/2020)	Deleterious
	S81L	New Zealand	Deleterious
ORF8	V62L	New Zealand	Neutral
	L84S	Multiple countries isolates[#]	Neutral
Nucleocapsid protein	L121H	China(hCoV-19/Shandong/LY003/2020)	Deleterious
	T148I	China(hCoV-19/Shenzhen/SZTH-004/2020)	Deleterious
	S193I	Wales	Deleterious
	S197L	Spain, Chile, Georgia	Neutral
	R203K	Multiple countries isolates[*]	Neutral
	G204R	Multiple countries isolates[*]	Neutral

	I292T	Switzerland	Neutral

$ Refer Supplementary table1, # Australia, Spain, Chile, USA, S.Korea, Georgia, China (hCoV-19/Hangzhou/ZJU-08/2020, hCoV-19/Beijing/235/2020, hCoV-19/Guangzhou/GZMU0047/2020, hCoV-19/Shandong/LY003/2020), India(hCoV-19/India/1-31/2020), Georgia, NewZealand, * Belgium, Brazil, Peru, Mexico, Nigeria, Vietnam, Switzerland, India(hCoV-19/India/c32/2020, hCoV-19/India/2020c32/2020, hCoV-19/India/c31/2020)

Supplementary Table

Supplementary Table 1: SARS Cov-2 virus used in the study. Isolates with P4715Lof RdRp and D614G of spike protein are indicated in bold

Location	Virus Name	Accession ID
China	hCoV-19/Beijing/235/2020	EPI_ISL_413521
	hCoV-19/Chongqing/ZX01/2020	EPI_ISL_408479
	hCoV-19/Guangzhou/GZMU0047/2020	EPI_ISL_414690
	hCoV-19/Hangzhou/ZJU-08/2020	EPI_ISL_416473
	hCoV-19/Hefei/2/2020	EPI_ISL_412026
	hCoV-19/HongKong/CUHK1/2020	EPI_ISL_416314
	hCoV-19/Jiangsu/JS03/2020	EPI_ISL_411953
	hCoV-19/Shandong/LY003/2020	EPI_ISL_414936
	hCoV-19/Shanghai/SH0067/2020	EPI_ISL_416370
	hCoV-19/Shenzhen/SZTH-004/2020	EPI_ISL_406595
	hCoV-19/Tianmen/HBCDC-HB-07/2020	EPI_ISL_412983
India	hCoV-19/India/1-27/2020	EPI_ISL_413522
	hCoV-19/India/1-31/2020	EPI_ISL_413523
	hCoV-19/India/763/2020	EPI_ISL_420543
	hCoV-19/India/2020763/2020	EPI_ISL_420544
	hCoV-19/India/770/2020	EPI_ISL_420545
	hCoV-19/India/2020770/2020	EPI_ISL_420546
	hCoV-19/India/772/2020	EPI_ISL_420547
	hCoV-19/India/2020772/2020	EPI_ISL_420548
	hCoV-19/India/773/2020	EPI_ISL_420549
	hCoV-19/India/2020773/2020	EPI_ISL_420550
	hCoV-19/India/777/2020	EPI_ISL_420551
	hCoV-19/India/2020777/2020	EPI_ISL_420552
	hCoV-19/India/781/2020	EPI_ISL_420553
	hCoV-19/India/2020781/2020	EPI_ISL_420554
	hCoV-19/India/c32/2020	EPI_ISL_420555
	hCoV-19/India/2020c32/2020	EPI_ISL_420556
	hCoV-19/India/1073/2020	EPI_ISL_421662
	hCoV-19/India/1093/2020	EPI_ISL_421663
	hCoV-19/India/1100/2020	EPI_ISL_421664

	hCoV-19/India/1104/2020	EPI_ISL_421665
	hCoV-19/India/1111/2020	EPI_ISL_421666
	hCoV-19/India/1115/2020	EPI_ISL_421667
	hCoV-19/India/1616/2020	EPI_ISL_421669
	hCoV-19/India/1617/2020	EPI_ISL_421670
	hCoV-19/India/1621/2020	EPI_ISL_421671
	hCoV-19/India/1644/2020	EPI_ISL_421672
	hCoV-19/India/1063/2020	EPI_ISL_424361
	hCoV-19/India/1135/2020	EPI_ISL_424362
	hCoV-19/India/1652/2020	EPI_ISL_424363
	hCoV-19/India/3118/2020	EPI_ISL_424364
	hCoV-19/India/3239/2020	EPI_ISL_424365
	hCoV-19/India/c31/2020	EPI_ISL_426179
Rest of Asia	hCoV-19/Cambodia/0012/2020	EPI_ISL_411902
	hCoV-19/Fujian/13/2020	EPI_ISL_411066
	hCoV-19/Georgia/Tb-390/2020	EPI_ISL_416477
	hCoV-19/Japan/SMU-0311S3/2020	EPI_ISL_416525
	hCoV-19/Kuwait/KU12/2020	EPI_ISL_416458
	hCoV-19/Pakistan/Gilgit1/2020	EPI_ISL_417444
	hCoV-19/Nepal/61/2020	EPI_ISL_410301
	hCoV-19/Saudi Arabia/SCDC-3324/2020	EPI_ISL_416522
	hCoV-19/Singapore/11/2020	EPI_ISL_410719
	hCoV-19/South Korea/KUMC05/2020	EPI_ISL_413515
	hCoV-19/Taiwan/CGMH-CGU-05/2020	EPI_ISL_415743
	hCoV-19/Nonthaburi/74/2020	EPI_ISL_403963
	hCoV-19/Vietnam/CM296/2020	EPI_ISL_416431
Europe	hCoV-19/Belgium/MTR-03021/2020	EPI_ISL_416467
	hCoV-19/Denmark/SSI-04/2020	EPI_ISL_416153
	hCoV-19/England/20100121007/2020	EPI_ISL_415141
	hCoV-19/Finlad/FIN-266/2020	EPI_ISL_414646
	hCoV-19/France/B2340/2020	EPI_ISL_416507

	hCoV-19/Germany/BavPat1/2020	EPI_ISL_406862
	hCoV-19/Hungary/mbl1/2020	EPI_ISL_416426
	hCoV-19/Italy/UiSR1/2020	EPI_ISL_413489
	hCoV-19/Luxembourg/Lux1/2020	EPI_ISL_413593
	hCoV-19/Netherlands/Flevoland_1/2020	EPI_ISL_415460
	hCoV-19/Poland/PL_P1/2020	EPI_ISL_416488
	hCoV-19/Russia/StPetersburg-3524/2020	EPI_ISL_415710
	hCoV-19/Scotland/EDB003/2020	EPI_ISL_415640
	hCoV-19/Spain/Valecia5/2020	EPI_ISL_416484
	hCoV-19/Sweden/01/2020	EPI_ISL_411951
	hCoV-19/Switzerland/GE1402/2020	EPI_ISL_415700
	hCoV-19/Wales/PHW32/2020	EPI_ISL_415920
Central America	**hCoV-19/Panama/328677/2020**	EPI_ISL_415152
North America	hCoV-19/USA/CA-CDPH-UC2/2020	EPI_ISL_413558
	hCoV-19/Canada/BC_02421/2020	EPI_ISL_415581
	hCoV19/Mexico/CDMXIDRE_01/2020	EPI_ISL_412972
	hCoV-19/USA/WA-S10/2020	EPI_ISL_416465
South America	**hCoV-19/Brazil/SPBR-03/2020**	EPI_ISL_414014
	hCoV-19/Chile/Santiago_op4d1/2020	EPI_ISL_415661
	hCoV-19/Peru/010/2020	EPI_ISL_415787
Africa	**hCoV19/Congo/K13/2020**	EPI_ISL_414647
	hCoV-19/Nigeria/Lagos01/2020	EPI_ISL_413550
Occeania	hCoV-19/New Zealand/20VR0275/2020	EPI_ISL_416538
	hCoV-19/Australia/VIC09/2020	EPI_ISL_416515
	hCoV-19/Australia/QLD09/2020	EPI_ISL_414414
	hCoV-19/Australia/NSW03/2020	EPI_ISL_408977

Supplementary Figure Legends

Fig. S1: (a)-(b) Structural changes in RdRp due to P4715L. (a) Change in secondary structure due to mutation wildtype exhibit turn and (b) P4715L mutant shows helix conformation. (c) –(d) Change of shape in epitope of RdRp. (c) wildtype RdRp (green) with the elliptical loop of epitope highlighted in

yellow. Pro4715 is displayed with ball and stick. (d) mutant RdRp (red) with the distorted elliptical loop of epitope highlighted in yellow. Leu4715 is displayed with ball and stick.

Fig. S2: Comparison of interatomic interactions in the wild-type and mutant form. Wild-type and mutant residues are marked in green represented in stick form. (a)-(b) Structural comparison due to P4715L RdRp. (c)-(d) Structural comparison due to D614G in spike protein.

Figures

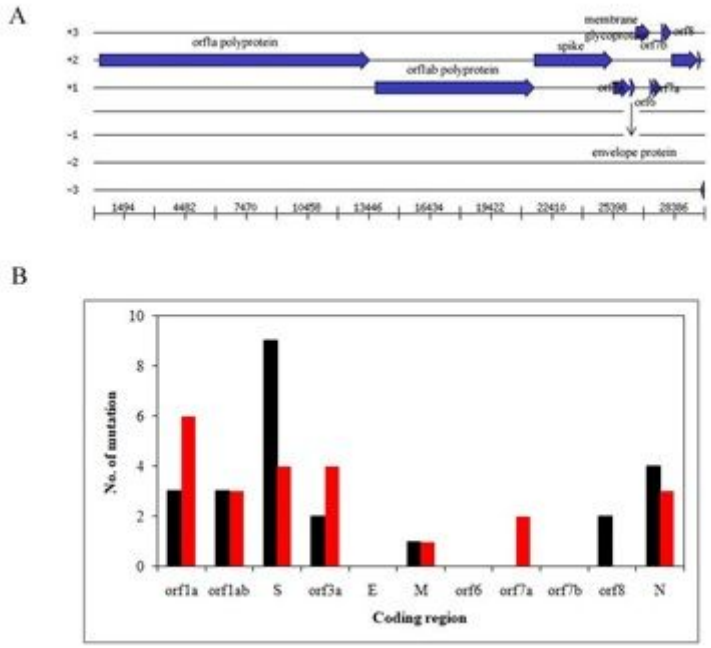

Figure 1

(a) Structure of the SARS-Cov-2 genome. (b) The relative occurrence of non-synonymous mutation across the genome of SARS-Cov-2. Black indicates the neutral effect of mutation and red are deleterious mutations.

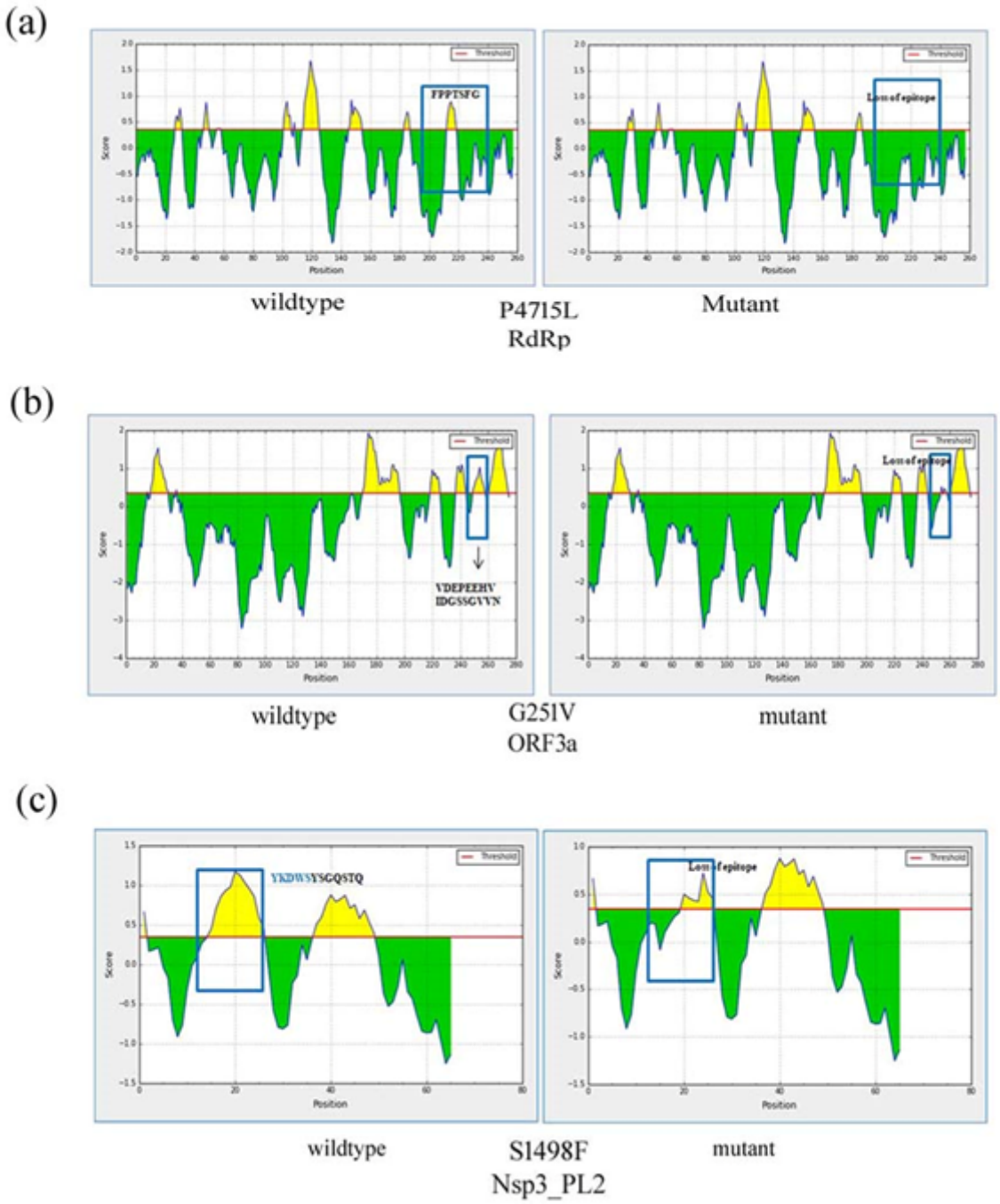

Figure 2

Epitope loss linked with non-synonymous mutations. The predicted score above the threshold level is the Yellow region showing epitope. (a) Effect of P4715L mutation in RdRp. (a) B-cell epitope in non-mutated RdRp (left) and P4715L mutant (right). It is linked with such epitope loss FPPTSFG from the site. (b) Epitope loss linked with ORF3a G251V. The B-cell epitope of wildtype ORF3a (left). G251V mutant (right) causes loss of DGSSGVV(250….256aa). (c) Epitope loss linked with S1498F in the papain-like protease domain of NSP3 in the ORF1a region. The B-cell epitope of wildtype sample (left), S1498F mutation causes the loss YKDWS (right).

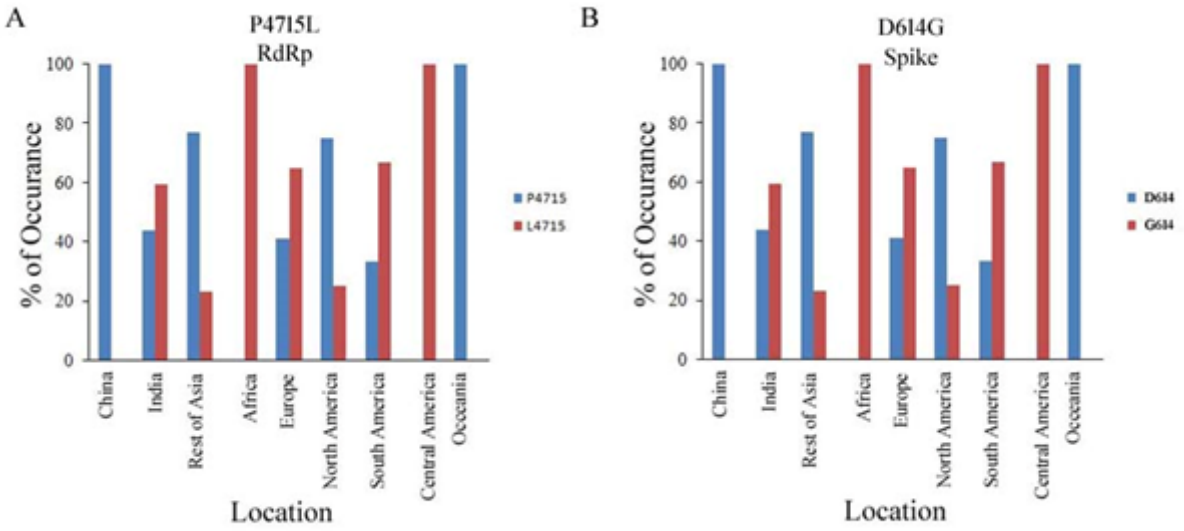

Figure 3

P4715L RdRp and D614G spike variants show co-occurrence. (a) The incidence (%) of occurrence of P4715 and L4715 of RdRp in various geographical locations. (b) The incidence (%) of occurrence of D614 and G614 of spike protein in various geographical locations. The distribution of these two unique mutant sites of vital SARS-Cov-2 proteins can be completely superimposed.

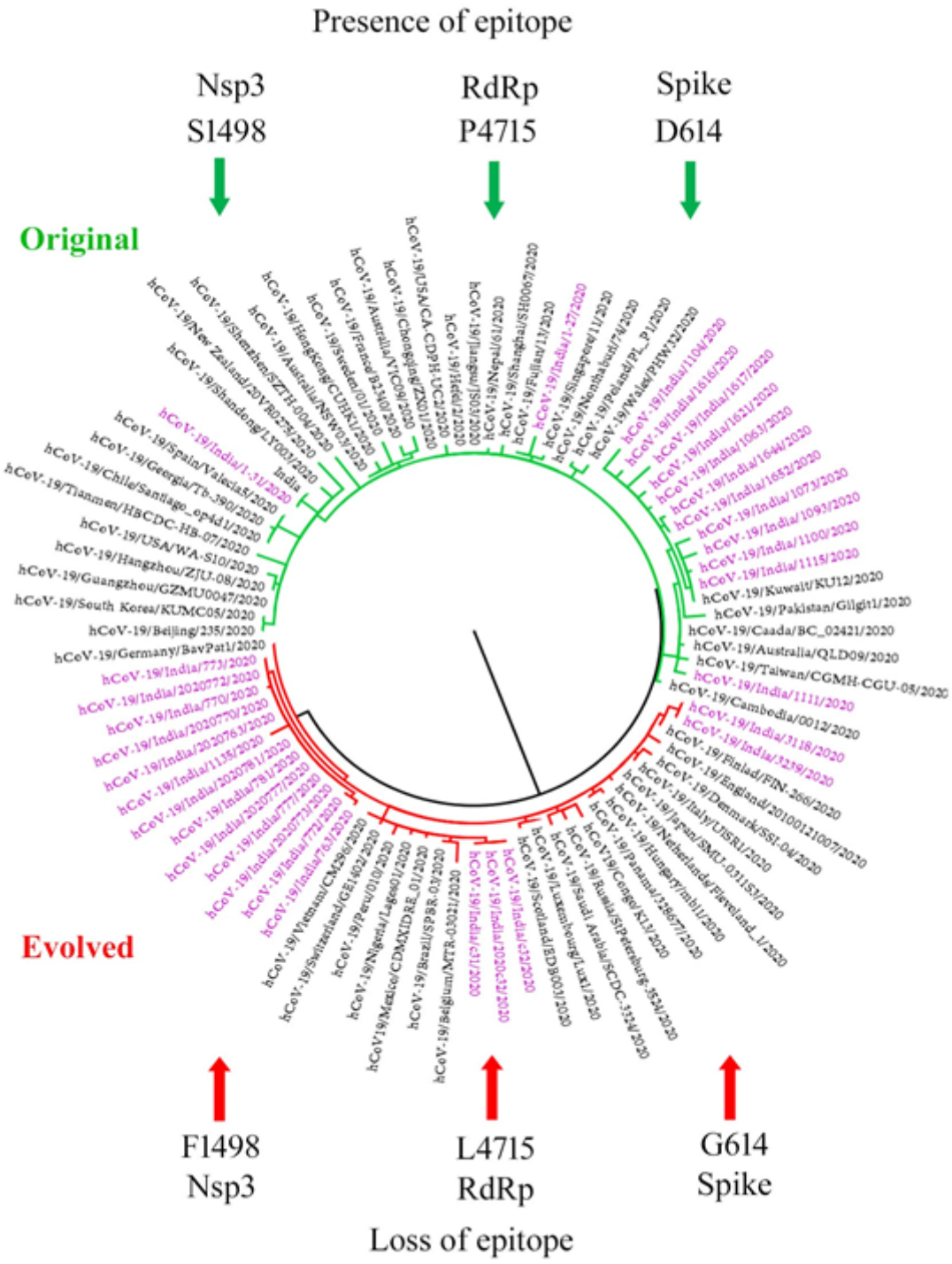

Figure 4

Phylogenomics with 87 SARS-Cov-2 viral samples across the world. The tree is distinctly divided into 2 clades: green shows the original (ancestral) form of the virus isolated from Dec, 2019 to Feb, 2020 and red clades are the evolved variant isolated after Feb, 2020 to April, 2020. Loss of epitope is found in the evolved variant due to non-synonymous mutation. Indian isolates are highlighted in purple belonging to both the green and red clades. The samples from Kerala, India are Wuhan representative (within green clade) and the rest are the variants from Italy (within red clade).

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- FigureS1ResSq.docx
- FigureS2ResSq.docx