ORIGINAL ARTICLE

A Study on Prevalence and Molecular Characterization of Spike Protein of SARS-COV-2 in Pakistan

Mahnoor Rizwan1, Abdul Ahad2, Sadia Aziz3, Nazir Ahmed Lone4, Tabasum Naz4, Muhammad Bilal Khan5, Laila Jafri4, Liaqat Ali4*

ABSTRACT

Objective: To examine the prevalence of antibody IgG/IgM and designing follow-up study of six months and to perform the molecular characterization and sequence analysis of spike protein of SARS-CoV-2.

Study Design: Prevalence study.

Place and Duration of the Study: The study was conducted in the Post-Graduate Lab of the National University of Medical Sciences (NUMS) Rawalpindi, Pakistan from January 2021 to June 2021.

Methods: 112 suspected patients with mild and severe symptoms were tested by PCR (polymerase chain reaction) and all were positive. Antibody profiling of vaccinated and non-vaccinated patients was performed by enzyme-linked immunosorbent assay (ELISA) and followed up for up to six months.

Results: Ten patients were identified as re-infected and were subsequently re-tested via PCR and DNA sequencing. The IgG/IgM of immunized patients was 11.87IU/mL and 3.87IU/mL in non-immunized individuals. Genomic analysis indicated that all the sequences belong to delta variant (21A and 21J sub-lineages). Mutations were observed in spike protein at I1169V, D950N in 21A and T951I and S254F in 21J sub-lineages.

Conclusion: This study revealed the significance of measuring antibody levels among vaccinated and unvaccinated individuals. Additionally, mutations in the spike gene of the delta variant of SARS-CoV2 can influence the pathogen’s mutation rate, leading to changes in its transmissibility and pathogenicity.

Keywords: Antibody Quantification, COVID-19, ELISA, Follow-Up, Immunity.

Introduction

One of the most pressing health concerns nowadays is the emergence of a new coronavirus strain, which belongs to the Coronaviridae family of RNA viruses, and has been causing illness and disruption worldwide. The very first case of pneumonia like disease was traced in Wuhan, China in 2019 December. Later on, it was discovered to be a novel strain of SARS-CoV-2 that was reported in other countries around the world and quickly became a pandemic. There are 4 naturally occurring species of the corona virus, classified into four genera: Alpha, Beta, Gamma, and Delta coronavirus. The human pathogenic coronaviruses i.e., HCoV229E, HCoVHKU1, HCoVNL63 and HCoVOC43 persistently cause infections in the human population and cause general symptoms of the common cold. The severe illness caused by this family of virus is acute respiratory syndrome (SARS) and the Middle East Respiratory Syndrome (MERS). The SARS CoV-2 is an enveloped virus having helical symmetry. It has a single stranded positive sense RNA with a layer of...
capsid around nucleus having genome size of approximately 30 kb. Generally, it has four structural proteins, Membrane protein (M), Nucleocapsid protein (N), Envelop protein (E) and Spike protein (S) and many other proteins (non-structural). SARS-CoV-2 has many transmission routes, it can be transmitted via short-distance droplet transmission, contact transmission aerosols in the enclosed space and urine, and mother-to-child transmission may also exist. In China, the virus has been isolated from fecal sample of a COVID-19 positive patient, which shows that it can survive in human feces as well. Virus replicates and mutates when making 24 copies. These mutations can lead to change in transmission and severity of illness causing a new variant of the same virus. SARS CoV-2 can infect many systems and organs of the human body. The sign and symptoms of COVID-19 vary in patients. When it comes to diagnosing the COVID-19 among patients it is difficult for a physician to identify because of different response of individual against the virus. The reason is the difference in genetics of the host determinants which fight against virus. Various immunological receptors and enzymes that react against corona virus are important to study because the medicine or vaccine developed can be easily tested and matched according to the host determinants. Current study focused onto investigates the prevalence of SARS-CoV-2 in capital region of Pakistan and link it with the various factors including age, gender, and immunity status.

**Methods**

**Sample collection**

The serum specimens are collected from a tertiary care hospital (Maroof International Hospital) situated in Islamabad, Pakistan. The antibody profiling was done in Post-Graduate Lab of the National University of Medical Sciences (NUMS) Rawalpindi, Pakistan from January 2021 to June 2021 after the approval of institutional review committee of the institute dated January 10, 2021 vide letter no (AUIC-LS3979). The sample patients selected were all COVID-19 positive (confirmed through PCR) and post COVID vaccinated. A written consent was taken from every patient to willingly participate in this research.

**Antibody profiling**

A total of 112 patients were targeted to check the antibody levels, in order to investigate the persistence of IgM/IgG and for follow up 50 patients were selected which were divided into vaccinated and non-vaccinated groups equally, in order to compare the antibody levels in both groups. Out of 50, 26 were COVID-19 positive. A written consent was taken from every participant. The antibody quantification was done through ELISA (Thermo Fisher Scientific - USA), using the electrochemiluminescence immunoassay (ECLIA) technique. The serum sample from whole blood specimen was separated using centrifuge. 3ul serum sample was placed inside the machine which detects the antibody levels and gives the result in 30 minutes.

Medical history of the patient and duration of injected vaccine was kept under consideration in order to check the antibodies titer and the start of decline phase in each patient against coronavirus. Samples were repeated after every one month to analyze the antibody titer during a period of six months. The follow up started from April 2021 and ended in September 2021. All the data collected was analyzed through statistical tool using Prism graph software. For genetic characterization following steps were taken and was performed at National University of Medical Sciences (NUMS) laboratory, Islamabad.

**RNA extraction and complementary DNA synthesis**

10 patients from follow up study were selected for genetic characterization. These patients were selected on the basis of re-infection. The age of the patients was between 20 to 40 years. Nasopharyngeal swab specimens were taken from individuals for RNA isolation purpose. Purification of viral RNA from nSARS-CoV2 infected serum sample was done using RNA Mag-MAXViral/Pathogen Nucleic Acid Isolation Kit, USA (mag max isolation kit) and for complementary DNA synthesis cDNA kit (thermo fisher, USA) was used. cDNA was quantified by using Nano drop technology.
**PCR and sequencing**

Gradient PCR was used to amplify the viral spike gene. Synthesized primer sequences were taken from national institute of health (NIH), Islamabad. Primers used had size ranging from 700-bp to 900-bp to cover the spike gene of SARS-CoV2. A total of 5 sets of primer were used to amplify the spike gene. After the successful amplification of spike gene, all the amplified PCR products were loaded onto on 1.2 % agarose gels depending upon the expected sizes of fragments and visualized under UV Tran's illuminator. Sanger sequencing technique was performed in order to analyze and detect any mutation of interest in spike gene of SARS-CoV2. The samples were sent to a commercial lab for sequencing purpose.

**Phylogenetic and mutational analysis**

The sequences received were in FASTA format. The sequences were uploaded on NCBI website for BLAST purpose to check the similarity with other SARS CoV-2 sequences. The first 43 sequences having 99% similarity were downloaded for further analysis using phylogenetic tree. Phylogenetic analysis was completed using the 46 accessible spike gene SARS-CoV2 attained from the NCBI SARS-CoV-2, sequences (https://www.ncbi.nlm.nih.gov/sars-cov-2/) were subjected to Multiple Sequence Alignment (M-S-A) via MAFFT online server. The M-S-A was afterward used to produce a Maximum Likelihood (M-L) phylogenetic tree via IQ-tree (http://www.iqtree.org/). For reference, Wuhan SARS-CoV2 (hCoV-19/Wuhan/IME-WH01/2019) sequence was utilized. The phylogenetic-tree was modified and pictured via Fig-tree software (http://tree.bio.ed.ac.uk/software/figtree/). Research sequences are highlighted in green color. For detection of mutations and comparison of the sample sequenced with similar sequences of same lineage clad next strain (https://clades.nextstrain.org) online software was used.

**Results**

The results of antibody prevalence during second wave of COVID-19 indicated that the COVID-19 has been mostly shown in age group of 30 and above. Table 1 indicates the antibody profile of patients according to age distribution.

<table>
<thead>
<tr>
<th>Gender</th>
<th>Antibody Average (IU/mL)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Male</td>
<td>12.98</td>
</tr>
<tr>
<td>Female</td>
<td>13.29</td>
</tr>
</tbody>
</table>

Majority of patients tested positive for COVID-19 were male Figure 1(c). Table 2 indicates the average of antibodies for six month follow up for each patient, indicating the infection positivity and immune system activation in response to the infection. 26 vaccinated patients had two shots of vaccine while 24 patients were non vaccinated.

The COVID-19 testing was done based on antibody quantification. If a person has antibody levels more than 0.8IU/mL, it indicates that the person is having the viral infection because the immune system has produced antibodies against the virus. People with age between 20 and 40 develop immunity against this virus without any vaccination because of a healthy immune system. The positivity rate of infection was higher among males as shown in Table 1. In adults and it is comparatively less while in old age people the rate was higher.

The antibody quantification pattern in Figure 1(a) is indicating that many of the patients were infected with COVID-19 and got recovered due to high antibody levels, as the antibodies do not rise during the first 14 days of the infection. The antibodies are produced after the 14 days of infection particularly.
**Table 2: Antibody titer among post-vaccination and non-vaccinated patients**

<table>
<thead>
<tr>
<th>S.NO</th>
<th>Vaccine dosage 1st dose</th>
<th>Booster dose</th>
<th>Non-vaccinated</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>18.42</td>
<td>62.5</td>
<td>3.89</td>
</tr>
<tr>
<td>2</td>
<td>46.27</td>
<td>47.1</td>
<td>74.5</td>
</tr>
<tr>
<td>3</td>
<td>71.84</td>
<td>151.2</td>
<td>4.5</td>
</tr>
<tr>
<td>4</td>
<td>48.9</td>
<td>83.14</td>
<td>48.2</td>
</tr>
<tr>
<td>5</td>
<td>83.12</td>
<td>102.4</td>
<td>56.05</td>
</tr>
<tr>
<td>6</td>
<td>49</td>
<td>57</td>
<td>28.32</td>
</tr>
<tr>
<td>7</td>
<td>6.36</td>
<td>18.2</td>
<td>0.8</td>
</tr>
<tr>
<td>8</td>
<td>15.39</td>
<td>10.05</td>
<td>2.29</td>
</tr>
<tr>
<td>9</td>
<td>134.4</td>
<td>121.3</td>
<td>0.05</td>
</tr>
<tr>
<td>10</td>
<td>50.6</td>
<td>66.8</td>
<td>13.4</td>
</tr>
<tr>
<td>11</td>
<td>23.02</td>
<td>56.1</td>
<td>24.46</td>
</tr>
<tr>
<td>12</td>
<td>9.94</td>
<td>22.3</td>
<td>15.33</td>
</tr>
<tr>
<td>13</td>
<td>35.6</td>
<td>82.3</td>
<td>6.26</td>
</tr>
<tr>
<td>14</td>
<td>1.4</td>
<td>74.2</td>
<td>1.3</td>
</tr>
<tr>
<td>15</td>
<td>50.05</td>
<td>51.5</td>
<td>7.9</td>
</tr>
<tr>
<td>16</td>
<td>40.29</td>
<td>72.1</td>
<td>1.11</td>
</tr>
<tr>
<td>17</td>
<td>44.1</td>
<td>48.9</td>
<td>22.0</td>
</tr>
<tr>
<td>18</td>
<td>22.1</td>
<td>30.9</td>
<td>10</td>
</tr>
<tr>
<td>19</td>
<td>71.2</td>
<td>77.1</td>
<td>10.1</td>
</tr>
<tr>
<td>20</td>
<td>55.9</td>
<td>82.8</td>
<td>18.0</td>
</tr>
<tr>
<td>21</td>
<td>67.2</td>
<td>81.2</td>
<td>151.2</td>
</tr>
<tr>
<td>22</td>
<td>22.3</td>
<td>34.6</td>
<td>4.3</td>
</tr>
<tr>
<td>23</td>
<td>79.3</td>
<td>121.2</td>
<td>1.9</td>
</tr>
<tr>
<td>24</td>
<td>181.5</td>
<td>189.7</td>
<td>10.9</td>
</tr>
<tr>
<td>25</td>
<td>34.9</td>
<td>55.9</td>
<td></td>
</tr>
<tr>
<td>26</td>
<td>88.09</td>
<td>92.3</td>
<td></td>
</tr>
</tbody>
</table>

*A titer of antibody levels after first dose of vaccination. *B titer of antibody levels after booster dose. *C antibody levels among non-vaccinated patients

---

**Fig 1 (a): Indicating antibody profiling with age groups ranging from 20 to 80**

The IgG antibody is produced after the infection while IgM is the first antibody produced during an infection indicating that a person has suffered from the infection recently.

**Fig 1 (b): indicating average of antibody levels each month**

**Antibodies levels against covid19**

---
The follow up data of 50 patients tested positive for COVID-19 as shown in figure 1(c). It indicates the elevation in antibodies each month indicating the infection positivity and immune system activation in response to the infection. Three of the patients died during the follow up study because of severe infection that led to lungs failure and cardiac arrest and it is clearly seen that no efficient number of antibodies was produced by their immune system and all three of them were non-vaccinated. While two of the female patients were pregnant and tested positive for COVID-19 during their pregnancy and both were 28 years old, one of them was vaccinated while the other was non-vaccinated. Their antibody quantification pattern indicated a raise each month showing a strong immune response against the infection. As both had equal elevations in their antibody levels so it means that during pregnancy immune system tends to produce more strong response against any infection because of the fetal development inside the body and the fetus also produces antibodies that aid in immune response. A few patients of middle age, non-vaccinated, did not develop effective number of antibodies and eventually got negative or non-significant after few months, indicating that either the person has got re-infected or has a low immunity against the infection. Most of the vaccinated patients had shown a clear elevated pattern of antibodies each month which indicates that vaccination is a supportive aid for immune system to fight against the particular infection and helps in protecting the body from re-infection. The average of antibody quantification in vaccinated patients is 11.87IU/mL, while in non-vaccinated it is 3.84IU/mL. People of young age without vaccination have shown an elevation in antibody levels each month indicates that they have a strong healthy immune system.

A clear elevation pattern can be seen in Figure 1(a) indicating the overall efficacy of immune system of the individual’s positive with COVID-19. This elevation pattern also indicates the development of herd immunity in the region which plays an important role for protection against COVID-19.

Three samples were successfully amplified and visualized on the gel documentation system after gel electrophoresis. The band size of the amplified gene (SARS-CoV-2 spike), was determined to be ~ 1000-bp. A 1kb ladder was used as a size reference, as shown in Figure 2. These samples were successfully sequenced commercially and submitted to NCBI for further analysis. The sequences submitted were given following accession numbers:

- Abasyn 01: ON115820
- Abasyn 02: ON115821
- Abasyn 03: ON115822

**Phylogenetic analyses**

All sequences were subjected to BLAST for similarity index. First 43 sequences were similar to the targeted sequences and they were selected for phylogenetic analysis as shown in Figure 3(a) and 3(b).

Phylogenetic analysis of 3 spike protein gene of SARS-CoV2 from the current study and other international strains of the virus similar to the study sequences were conducted. The mean of pair wise distance between USA and our sequences were indicating phylogenetic relatedness between the genomes figure 3(a). The SARS-CoV-2 lineage diversity has been inspected in the current study. The study has shown the Delta (n=03) variant of SARS-CoV-2. In delta variant 2 clads have been identified i.e., 21A and 21J.
Mutational analysis

The mutational analysis was done on clad next strain software. The mutations in each sequence were highlighted along with the nucleotide positions. The sequence under accession number ON115820, showed mutations on 8 positions in nucleotides as shown in Figure 4(b). The sequence under accession number ON115821, showed mutations on 10 positions in nucleotides, as shown in Figure 4(c). The sequence under accession number ON115822, showed mutations on 7 positions in nucleotides, as shown in Figure 4(a). These 3 sequences of interest were compared with the 4 other sequences that showed similarity in phylogenetic analysis. With the use of next clade software, sequences being analyzed for mutational purpose were subjected to phylogenetic analysis.

The mutational analysis results indicated that the sequences of clade 21A had the mutations I1169V and D950N. The sequences of clade 21J had mutations T951I and S254F. The mutations indicated that threonine, serine and isoleucine all are uncharged amino acids and got substituted by uncharged isoleucine, phenylalanine and valine respectively. This shows that the binding affinity of RBD did not change due to these mutations as all were uncharged amino acids. Herd immunity plays vital role in the infection cycle and can cause the virus to undergo non-significant mutations. This indicates that mutations can occur differently in different regions of the world which affects the pathogenicity and transmissibility of the virus.

The lineages AY103 viruses of delta variant was identified in America while lineage b.1.617.2 was identified in India. Our sequence of interest was of lineage B.1.617.2, while similar sequences of clade 21 were of American variant while clade 21A belong to Indian isolated delta variant. It indicates that the delta variant circulating in Pakistan is most probably the Indian isolated strain with different mutations.
which can be the reason of its less severity and transmissibility.

**Discussion**

The prevalence of delta variant in Islamabad was high as shown in results but the severity of infection was low in Pakistan as compared to other countries that suffered from delta variant. The mortality rate of delta variant was very high in India and other regions of the world in all age groups except infants and children of age 10 and above. According to a study seroprevalence of SARS-CoV-2 among Africans is more than the other ethnic groups of the world. This can be due to the less development of herd immunity among the community or difference in lifestyle of people living in Africa.  

According to the statistics of antibody quantification performed in a tertiary care hospital of Islamabad, the antibody levels increase with passage of time in order to build protection against the particular infection and prevent the body from re-infection. COVID-19 has been a threat across the globe causing many damages; the mutational rate of this virus is high causing more pathogenicity and transmission of the infection. The antibody quantification shown in Table 1 indicates that in youth mostly the immune system of the infected person is healthy, developing immunity rapidly against the infection. Most of the patients have age of 20 to 50 and half of them were vaccinated in order to observe the antibody pattern. The vaccinated patients have a greater average of antibody quantification as compared to the non-vaccinated patients which means that vaccination is an effective and supportive aid against COVID-19. However, antibody testing is not a reliable method to indicate whether a person is infected with COVID-19 or not but it can give a good idea about the immunity status of the infected patients.  

The positivity rate of COVID-19 has been observed higher in males as compared to females in the current study. This similar pattern has been observed in various other studies of seroprevalence in other regions of the world. Male population was affected more by COVID-19 than females because of the higher rate of social interactions among males as compared to females. As COVID-19 can only be prevented through quarantine and social distancing which is seen less in male population because of the activities of males outside the house like jobs, sports etc.  

The age group of 20 to 60 years old has been mostly infected with COVID-19 with different recovery time depending upon the immunity status of the individual. The results of the current study indicate the fast recovery of patients without hospitalization due to healthy immunity and diet. While in other regions of the world hospitalization rate of adults
was surprisingly high due to higher rate of infection and its severity. Mostly the positivity rate of COVID-19 increased initially due to asymptomatic patients.

The recovery of asymptomatic patients was seen fast and easy while the transmissibility of infection due to it was high.
During the first year of COVID-19 pandemic in the world, the SARS-CoV-2 constituted few mutations with the circulating viruses being closely related to Wuhan-HU-1 strain. The D614G substitution was the first mutation found in the SARS-CoV-2 with worldwide prevalence and now the D614G is one of the commonest mutations found in all the circulating lineages and sub-lineages of SARS-CoV-2. The severity rate of this variant has been less in Pakistan as compared to other regions because of mutational changes occurring in spike protein of SARS-CoV2. This variant spread to 179 countries causing different infection patterns due to different mutations observed in its spike protein. The current study indicated one new mutation occurring at the sequences of clade 21A had the mutations I1169V and D950N. The sequences of clade 21J had mutations T951I and S254F. while the early mutations observed in this variant were Y145H and A222V substitution.

Current study had shown two sub clades of delta variant which were first reported in India and USA. Upon comparative analysis with other sequences of spike gene of delta variant it was observed those two new mutations were found which were not present in the clades isolated from other regions. This justifies those regional mutations can occur in the virus which affects its infection rate and transmissibility rate.

The mutations were found to be linked with amplified transmissibility and binding affinity of virus with ACE2 compared to D614 viral strains. The subtleties of pandemic have transformed worldwide after September, 2020 with the appearance of novel viral strains containing huge number of mutations in the genome. These novel strains have been categorized as alpha, beta, gamma, and delta. In spite of these main lineages of the SARS-CoV2, there are many other lineages and clades of virus that have arisen, which altered the worldwide pattern of pandemic. The variants help virus in survival by increasing transmissibility, pathogenicity and immune escape by deactivating antibodies. The genomic investigation has discovered the prevalence of delta variant during fourth wave of pandemic. A few novel mutations have been detected, that might be the cause of reduced severity of the delta variant circulating in Pakistan. While due to a lesser number of samples complete prevalence and mutational analysis could not be identified but the obtainable data gives a good idea of the novel mutations that panels the viral pathogenicity.

**Conclusion**

The prevalence of antibody quantification among vaccinated and non-vaccinated patients indicated the significance of herd immunity and vaccination as a supportive aid in fighting against COVID-19. Age factor and vaccination status were found to be key players in determining the extent of immune response of the body and severity of infection. The mutations expressed in spike gene of delta variant shows that the region may also affect the mutational rate of the pathogen and may cause a change in its pathogenicity and transmissibility.

**Authors Contribution**

MR: Data collection, data analysis, results and interpretation, manuscript writing and proof reading
AH: Data collection, data analysis, results and interpretation
SA: Data collection, data analysis, results and interpretation, manuscript writing and proof reading
NAL: Data analysis, results and interpretation, manuscript writing and proof reading
TN: Data analysis, results and interpretation, manuscript writing and proof reading
MBK: Data analysis, results and interpretation, manuscript writing and proof reading
LJ: Data analysis, results and interpretation, manuscript writing and proof reading
LA: Idea conception, study designing, data analysis, results and interpretation, manuscript writing and proof reading

**REFERENCES**


