Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
Identification of E484K and other novel SARS-COV-2 variants from the Kingdom of Bahrain

Khalid Mubarak Bindayna a,*, Abdel Halim Abdel Fattah Salem Deifalla b, Hicham Ezzat Mohamed Mokbel a

a Microbiology, Immunology and Infectious Diseases, Bahrain
b Anatomy Department, College of Medicine and Medical Sciences, Arabian Gulf University, Bahrain

ABSTRACT

The challenges imposed by the ongoing outbreak of severe acute respiratory syndrome coronavirus-2 affects every aspect of our modern world, ranging from our health to our socio-economic needs. Our existence highly depends on the vaccine’s availability, which demands in-depth research of the available strains and their mutations. In this work, we have analyzed all the available SARS-COV2 genomes isolated from the Kingdom of Bahrain in terms of their variance and origin analysis. We have predicted various known and unique mutations in the SARS-COV2 isolated from Bahrain. The complexity of the phylogenetic tree and dot plot representation of the strains mentioned above with other isolates of Asia indicates the versatility and multiple origins of Bahrain’s SARS-COV2 isolates. We have also identified two high impact spike mutations from these strains which increase the virulence of SARS-COV2. Our research could have a high impact on vaccine development and distinguishes the source of SARS-COV2 in the Kingdom of Bahrain.

1. Introduction

Severe Acute Respiratory Syndrome CoV-2 or SARS-CoV-2 is the most dangerous threat to humankind across the globe. This disease became a pandemic in early 2020 and affects various countries worldwide. It not only affects the human race by its disastrous impact on human health but also through its devastating influence on every socio-economical aspect of the modern world. It was first reported in China at the city of Wuhan in December 2019. The novel coronavirus, the causative agent of COVID-19, can be transmitted from man to man through body secretions. SARS-CoV-2 is an enveloped and single-stranded positive-sense RNA containing virus which is closely related to the coronavirus that has been isolated from SARS outbreak in 2003. Though there are some shreds of evidence for the spillover infection of the zoonotic origin of SARS-CoV-2, some controversies also exist regarding its laboratory origin (1).

In the Kingdom of Bahrain, the very first case was detected on February 21, 2020 in a school bus driver who had a travel record to Iran and Dubai. Within three months, the total number of cases of SARS-CoV-2 in the Kingdom of Bahrain reached $10^3$. Though there was a reduction in the increasing rate of new cases after six months, it again started increasing at the beginning of September. This could be counted as the second wave, which also has started decreasing at the end of November.

This work aims to analyse up to date 150 SARS-CoV-2 genomes that were isolated from the Kingdom of Bahrain. With state of the art comparative genomic approach, we have analyzed the mutations present in those strains which will help in the vaccine design. We have also compared all the genomes with twenty different SARS-CoV-2 genomes that were isolated from various parts of Asia, to study the proper origin of SARS-CoV-2 in the Kingdom of Bahrain.

2. Method

2.1. Sample selection

We have fetched all the genomes of SARS-COV2 from the Global Initiative on Sharing All Influenza Data (GISAID) server [2]. Total 150 genomes with the origin of Bahrain and another 20 genomes with other Asian roots have been identified. The other 20 locations under consideration are Zahedan, Wuhan, Tehran, Semnan, Riyadh, Qom, Qatif, Qatar, Muscat, Makkah, Madinah, Lebanon, Kuwait, Jordan, Jerusalem, Jeddah, Islamabad, Iraq, Dhaka and Delhi. We have only collected a
2. Multiple sequence alignment

Multiple sequence alignment has been performed for two groups of the sample. One group contains only the SARS-COV2 sequences isolated from Bahrain and the other group contains all the genomes from Bahrain and previously mentioned twenty different locations. We performed MSA for both the groups for SNP detection in the first group and phylogenetic tree construction in the second group. Here, we have used the MUSCLE (Multiple Sequence Comparison by Log-Expectation) algorithm to align the sequences [3]. The MUSCLE algorithm is based on two distance measures: k-mer distance and Kimura distance for each of the two sequences. Before multiple alignments, scores for pairwise alignment have been defined. Initially, a binary tree has been created which is followed by accurate multiple alignments in the refinement stage through calculations of time and space complexities.

2.3. Variant and mutations analysis

We have used hCoV-19/Wuhan/WIV04/2019 strain (isolated from Wuhan, China) as a reference for variant and mutation analysis of 150 SARS-COV2 genomes isolated from the Kingdom of Bahrain. We have used the CoVsurver mutant analysis tool from the GISAID server. This tool identifies specific mutations of SARS-COV2 along with various structural and functional proteins. Similarly, for the characterization of Single Nucleotide Polymorphisms (SNPs), we have used SNiPlay tools available at (http://sniplay.cirad.fr/). This tool helps us in the determination of Single Nucleotide Polymorphisms (SNPs) and insertion/deletion (indels) along the whole genome length. Both the tools accept aligned FASTA sequences.

2.4. Phylogenetic tree construction

Phylogenetic tree construction is required for the detection of similarities in SARS-COV2 genomes isolated from various places. It is also helpful to detect the close relative or source of the 150 different SARS-COV2 genomes isolated from the Kingdom of Bahrain. For phylogenetic tree reconstruction, we have used the PhyML algorithm which is based on the maximum likelihood method [4]. It allows very large datasets in multiple sequence aligned format. This method follows the hill-climbing algorithm which simultaneously adjusts tree topology and branch lengths. It starts with a basic tree buildup via the fast distance-based method and with every iteration, it improves the likelihood of the tree. It is a very fast method which reaches optima within a few iterations.

2.5. Dot plot assay

The Dot plots are used to compare large sequence sets to get an insight into the similarity overview. A distinct set of sequences will produce a diagonal straight line. We have used D-genies [5] which is under GNU General Public License (GPL). Here, similar sequences will produce a scatter plot as most of the sequences will be matched with each other. A large number of samples could be plotted with a multiple sequence aligned format as an input file. We have used this plot to estimate the overall 170 previously mentioned SARS-COV2 sequences to check their similarities and relatedness.

![Fig. 1. Mutations on the spike glycoprotein in the unbound state (PDB id: 6ACC) and bound state (PDB id: 6acj) with host ACE2 receptor (green) in Fig. 1 (A) and (B) respectively. (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)](image-url)
3. Results

3.1. Mutation analysis

We have compared a total 150 genomes that have been isolated from the Kingdom of Bahrain with the reference genome from Wuhan strain (hCoV-19/Wuhan/WIV04/2019) which is the official reference sequence employed by GISAID. We have analyzed all the mutations in the 150 genomes compared to the reference strain. A detailed list of unique and known mutations is given in Supplementary Information 1. The unique mutations are not reported previously. Here, we have listed down eight strains which have ≥5 unique mutations in their genomes in Table 1.

We have also summarized the positions of these mutations on the spike glycoprotein in the unbound state (PDB id: 6ACC) and bound state (PDB id: 6ACJ) with host ACE2 receptor (green) in Fig. 1 (A) and (B) respectively [6,7].

Mutations on the spike glycoprotein that were previously identified have been colour coded in the given structure. H49Y[blue], W64L, A67S, G75S(77), T76I(77)[magenta], R78 M, R102I[blue], V127F, L141del[sky], G142del[sky], V143del[sky], Y144del[143][sky], E154G, E156D, M177I, S255F(257)[blue], I358V[yellow], L452 M [yellow], S459F[yellow], E484Q[yellow], E484K[yellow], L513F, D614G[blue], V622I, Q675H(674)[blue], S689I(692), A871S, I909V, L1063F, E1092A and H1101Y. Stain hCoV-19/Bahrain/BAH-24/2020|EPI_ISL_483638|2020-04-07 is identified to possess the maximum number of unique mutations. The analyzed genome variations in terms of SNP’s for these 150 genomes are given in Supplementary Information 2. Among these mutations, E484K and D614G are recently noted as significant mutations due to their effect on SARS-CoV-2 virulence. Total 6 isolates from the Kingdom of Bahrain possess E484K mutation at 23012 positions of their genome as represented by the multiple sequence alignment and Single Nucleotide Polymorphisms analysis in Fig. 2. These six isolates also possess D614G mutations. All the unique and known mutations along with the isolate’s name are represented in Table 2.

E484K mutation on the spike protein causes SARS-CoV-2 to escape from the neutralizing effect of the human immune systems mediated by antibodies [8]. On the other hand, D614G mutation on the spike protein has high infectivity in human cells having ACE2 receptors [9,10]. D614G mutation situated exactly at the receptor-binding domain of the spike protein. So due to this mutation, the affinity of the ACE2 receptor

![Fig. 2. Identification of E484K mutation on the genome of the six isolates as represented by the multiple sequence alignment and Single Nucleotide Polymorphisms analysis.](image)

| Query | Unique Mutation List | Existing Mutation List |
|-------|----------------------|------------------------|
| EPI_ISL_632260 | NSP5_K236 N, Spike_E156D | NSP3_Q1884H,NSP3_T73L, Spike_D614G,Spike_E484K, NS3_Q57H |
| EPI_ISL_632278 | NSP5_K236 N, Spike_E156D | NSP3_Q1884H,NSP3_T73L, Spike_D614G,Spike_E484K, NS3_Q57H |
| EPI_ISL_678270 | NSP3_A99V, Spike_E156D | NSP3_Q1884H,NSP3_I385T, Spike_D614G,Spike_E484K, NS3_Q57H,N,S193I |
| EPI_ISL_682320 | NSP3_A99V, Spike_E156D | NSP3_Q1884H,NSP3_I385T, Spike_D614G,Spike_E484K, NS3_Q57H,N,S193I |
| EPI_ISL_684032 | NSP8_V160I, Spike_V1164A, Spike_E156D | NSP3_Q1884H,NSP3_I385T, Spike_D614G,Spike_E484K, NS3_Q57H,N,S193I |
| EPI_ISL_684033 | NSP8_V160I, Spike_V1164A, Spike_E156D | NSP3_Q1884H,NSP3_I385T, Spike_D614G,Spike_E484K, NS3_Q57H,N,S193I |

![Fig. 3. Interaction of SARS CoV-2 Spike protein at E484 position (green) with human IgG heavy chain domain (brown). (For interpretation of the references to colour in this figure legend, the reader is referred to the Web version of this article.)](image)
increases. It was reported that infectivity increases by $>1/2 \log_{10}$ which is measured through cell fusion assay [9].

Now to understand the actual effect of the D614G and E484K mutations in these six Bahraini isolates, we have performed in silico docking assay of wild type SARS CoV-2 Spike protein (PDB id: 6ZB4) with human IgG antibody heavy chain (PDB id: 7CM4_2 and 6ZER_3). Fig. 3 indicating the interaction of E484K with IgG antibody.

Now when we have inserted both the mutations (E484K and D614G), the binding affinity has been significantly altered. We have studied the affinity using PANDA (Predict chAnge iN binDing Affinity) [11] of both the spike glycoprotein and IgG heavy chain. According to the algorithm, the affinity has been decreased by $-0.476$ kcal/mol. Such alteration of affinity is also experimentally observed in other SARS-COV2 stains with E484K mutation which could help SARS-COV2 to evade the neutralizing antibody [12].

We have also performed the background check of the six isolates that possess both the E484K and D614G mutation. Using the PANGO lineage web server, we have found that all the isolates belong to B.1.281 lineage. Besides Bahrain, this lineage has been found in Denmark. So both the mutations might have the same origin.

3.2. Phylogenetic tree analysis

We have reconstructed phylogenetic trees separately for the SARS-COV2 sequences isolated from the Kingdom of Bahrain and all 170 sequences from all over Asia. Here, first of all, our main aim was to identify the similarities between SARS-COV2 sequences isolated from the Kingdom of Bahrain. A view of the phylogenetic tree only for Bahrain sequences has been presented in Fig. 4. It can be observed from Fig. 4 that there is a huge diversity in the genomic sequences isolated from various places of Bahrain. This could definitely increase our curiosity for finding the actual origin of SARS-COV2 in Bahrain. So we constructed the second tree which possesses all the 170 sequences from Bahrain as well as the sequences from various places of Asia. The phylogenetic tree for all the 170 sequences is represented in Fig. 5.

From Fig. 4, it seems that the origin of SARS-COV2 in Bahrain is not only from one place. There is huge diversity in the phylogenetic tree where we can see that 37 sequences of SARS-COV2 isolated from Bahrain is very much different from other sequences. Other 113 sequences have a different origin. Some of the sequences have similarities with Qatif where some other sequences have similarities with the sequence isolated from Iran, Qatar and Israel. Such huge diversity leads us to perform a dot plot assay for the overall understanding of all the 170 sequences.
3.3. Dot plot assay

We have placed all 170 sequences in a specific manner both in the X and the Y-axis of a graph, and study the similarity matrix between them. The dot plot is represented in Fig. 6.

The scattered nature of the dot plot in Fig. 6 clearly indicates that all the 170 genomes are closely related to each other. If there is a single origin of the 150 SARS-COV2 genomes isolated from Bahrain, then the plot looks like a diagonal straight line. So we can conclude that the origin of SARS-COV2 in Bahrain is not only from one place.

4. Discussion

In this paper we have represented a detailed account of a total 150 SARS-COV2 genomes that have been isolated from the Kingdom of Bahrain. We have analyzed the mutations in all the strains in reference to the hCoV-19/Wuhan/WIV04/2019 Wuhan strain. We have identified various known and unknown mutations from these 150 strains. We saw that hCoV-19/Bahrain/BAH-24/2020|EPI_ISL_483638|2020-04-07 possesses a maximum number of unique mutations. We have found two significant mutations (E484K and D614G) on the spike protein of Bahrain isolates. These mutations cause decrease in antibody affinity towards SARS-COV2 and increase of binding affinity towards ACE2 receptor respectively. A detailed account of all these mutations will be helpful for designing the vaccine against SARS-COV2. We have also studied the origin of SARS-COV2 in the Kingdom of Bahrain. We found that there is a huge diversity in the 150 genomes. This indicates that there could be multiple sources of SARS-COV2 in the Kingdom of Bahrain.

CRediT authorship contribution statement

Khalid Mubarak Bindayna: Conceptualization, Methodology, Writing - original draft. Abdel Halim Abdel Fattah Salem Deifalla: Data curation, Visualization, Software, Investigation. Hicham Ezzat Mohamed Mokbel: Supervision, Validation, Writing - review & editing.
Declaration of competing interest

The authors of this paper declare no conflict of interests.

Acknowledgment

We thank the ministry of heath Bahrain and puplic health labratoty for for submitting the virus sequencing into GISAID.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.micpath.2021.104955.

References

[1] K.G. Andersen, A. Rambaut, W.I. Lipkin, E.C. Holmes, R.F. Garry, The proximal origin of SARS-CoV-2, Nat. Med. 26 (4) (2020) 450–452, https://doi.org/10.1038/s41591-020-0820-9.
[2] Y. Shu, J. McCauley, GISAID: Global initiative on sharing all influenza data—from vision to reality, Euro Surveill. 22 (13) (2017) 30494.
[3] R.C. Edgar, MUSCLE: multiple sequence alignment with high accuracy and high throughput, Nucleic Acids Res. 32 (5) (2004) 1792–1797.
[4] S. Guindon, O. Gascuel, A simple, fast, and accurate algorithm to estimate large phylogenies by maximum likelihood, Syst. Biol. 52 (5) (2003 Oct) 696–704, https://doi.org/10.1080/10635150390235528. PMID: 14530136.
[5] F. Cabanettes, C. Klopp, D-GENIES: dot plot large genomes in an interactive, efficient and simple way, PeerJ 6 (2018) e4958.
[6] J.T. Ortega, M.L. Serrano, F.H. Pujol, H.R. Rangel, Role of changes in SARS-CoV-2 spike protein in the interaction with the human ACE2 receptor: an in silico analysis, EXCLI journal 19 (2020) 410.
[7] C. Zhang, W. Zheng, X. Huang, E.W. Bell, X. Zhou, Y. Zhang, Protein structure and sequence reanalysis of 2019-nCoV genome refutes snakes as its intermediate host and the unique similarity between its spike protein insertions and HIV-1, J. Proteome Res. 19 (4) (2020) 1351–1360.
[8] E. Mahase, Covid-19: what New Variants Are Emerging and How Are They Being Investigated?, 2021.
[9] J. Ogawa, W. Zhu, N. Tonna, O. Singer, T. Hunter, A.L. Ryan, G.M. Pao, The D614G mutation in the SARS-COV2 Spike protein increases infectivity in an ACE2 receptor-dependent manner, Biorxiv (2020), https://doi.org/10.1101/2020.07.21.214932.
[10] S. Kim, J.H. Lee, S. Lee, S. Shim, T.T. Nguyen, J. Hwang, S. Kim, The progression of sars coronavirus 2 (SARS-COV2): mutation in the receptor-binding domain of spike gene, Immune Network 20 (5) (2020).
[11] W.A. Abbasi, S.A. Abbasi, S. Andleeb, PANDA: Predicting the Change in Proteins Binding Affinity upon Mutations Using Sequence Information, 2020 Sep 16 arXiv preprint arXiv:2009.08869.
[12] S. Jangra, C. Ye, R. Rathnasighe, D. Stadlbauer, F. Krammer, V. Simon, L. Martinez-Sobrido, A. Garcia-Sastre, M. Schotsaert, PVI study group, The E484K mutation in the SARS-CoV2 spike protein reduces but does not abolish neutralizing activity of human convalescent and post-vaccination sera, medRxiv (2021 Jan 1).

Fig. 6. A dot plot assay of all the 170 sequences isolated from various regions of Asia.