SARS-CoV-2 genome variants epidemiology surveillance in Ethiopia and dynamic mutational changes of structural S Spike protein through computational analysis

Ayele Abaysew Flifl (✉ betre.auroun@gmail.com)

Technology and Innovation Institute

Rita Majumdar Singh

Sharda University

Yohannes Sitotaw

Ethiopian Biotechnology Institute

Tesfaye Adisu Tarekegn

Near East University

Research Article

Abstract

Objective: This study aims to identify the variants of SARS-CoV-2 that were circulating in Ethiopia and spot dynamic mutational changes of spike antigenicity based on genome data analysis to put forward preventative measurement against pandemic.

Results: The SARS-CoV-2 genomes from Ethiopia were confirmed to be evolutionary related to RaTG13 and SL- bat coronavirus and Spike receptor sites were conserved. The clade distribution of the genome was reflected as GH, GR and other O and intended for new variants. 3 female samples were detected as Variants of Interest VUI202012/01GRY B.1.1.7 which Pango linage B.1.1.7 was originated from the UK. Despite 21 notable mutations, 71% D614G, 28% D614X, 35% N501Y and 21% NSP5-S284G mutation occurred predominantly in our genome samples and could have antigenicity and infectivity effects. Mutation on N440K was perceived as antigenic-drift in a sample and potency resist SER-52 antibody neutralization and vaccine escape.

Introduction

Novel coronavirus is becoming a major threat to the world and killing around 4 million people out of more than 180 million confirmed cases as of June, 2020 [1]. The first virus entered Ethiopia by a Japanese passenger who was flying to the capital city Addis Ababa by airplane in March 2020, according to the Ministry of Health report. Nowadays, the number of people infected with the virus continues to increase. 4,220 fatalities cases are registered and 273,175 total confirmed cases are enumerated and 1,894,717 vaccine doses have been administered in the country as of June 2021 [1, 2].

Since the virus entered and distributed in Ethiopia, molecular diagnosis techniques such as SARS-CoV-2 Antibody-IgG test were performed in Ethiopia [3]. Genomic study is not yet. Though a number of whole genome sequences of SARS-CoV-2 Ethiopian were sequenced and deposited in GISAD databases, no previous studies have been conducted in terms of genomics characterization and could not discover the variants in the country. It is very important to know the structural mutations and strains that are circulating in Ethiopia to ensure information about the potential of variants and promising indicators so as to control dynamic mutations. Presently, SARS-CoV-2 with the D614G and N501Y mutation has become the most frequently detected globally. As a result, a patient with D614G mutation has more viral loads and it is difficult to detect the virus easily using usual habitual techniques [4]. It was examined to neutralize mAbs test on those who were suffering from SARS-CoV-2 variants B.1.1.7 and lastly taken a sera from recipients of the oxford AstraZenic, Pfizer-Biotech vaccine [5]. Variants of B.1.351 N501Y also confirmed its resistance to therapeutically monoclonal antibodies [6]. Hence, the country like Ethiopia is likely poor and very difficult to overcome those dynamic mutations without genomic surveillance checkpoints.

In this study, we constructed evolutionary relatedness of the SARS-CoV-2 genome and detected variant concerns that were circulating in Ethiopia. Conserved domain of (RBD) protein sites and mutational dynamics of glycoprotein of structural S spike protein were discovered and assessed its potential against mAb-antibody. Our study provides full evidence about the variants and dynamic mutational changes that occurred in Ethiopia and used as the baseline research for tracking epidemiology of the virus.

Materials And Methods

(N=14) whole genome sequence of SARS-CoV-2 which was sequenced from Ethiopia. COVID-19 positives were retrieved from GISAID datasets and performed MSA through blastn techniques (E=0) and purposeful (n=67) cross regional complete genomes were booked for further analysis. MSA was generated through CLUSTALX software and the Phylogenetic tree was constructed using MEGAX (V. 10.1.7) software [7].

Variants, clade distribution, structural spike and none structural protein mutational changes and other epidemiology analysis was done through GISAID CoVsurver platforms https://www.gisaid.org/. The Protein amino acid sequence was analyzed using Jalview (version 2.10.0.) [8] and Conserved Domain analysis platforms. Finally, interaction between Spike receptor binding sites [PDB: 6acj] and ACE-2 as well flanked sites [PDB: 7k8t] and mAbs were analyzed through the pyMol (V 2.3.) visualizer [9].

Results

Phylogenetic Tree Analysis Related to Precursor

To study the genetic evolutionary relationship of the virus of that entered into Ethiopia, the evolutionary phylogenic tree analysis were conducted using MEGA using neighbor-joining statistical approaches and Poisson model along phylogeny test of 1000 bootstrap replication methods and found out that bat coronavirus RaTG13: MN996532.1 which were previously isolated from Yunnan province was considered as the nearest ancestor family to all Ethiopian SARS-CoV-2 followed by pangolin bat-SL-CoVZC45: MG772933.1 and bat-SL-CoVZXC21: MG772934.1 respectively. However, (MERS: KT806053.1 and MK564474.1) coronavirus was disclosed as a genetic distance as inferred on **Figure 1A.** This finding provisionally took a scientific evidence as per [10, 11], reported *Rhinolophus affinis* was consider a natural reservoir of the coronavirus and supposed to a source of future epidemic.

Nevertheless 11 patients SARS-CoV-2 genome samples from Ethiopia were grouped as in same genetic distance as it was marked in red color of the tree, it was reflected that sub family member of EPI ISL 1897884, EPI ISL 1899281 and EPI ISL 1899047 were clustered and genotype association to SARS-CoV-2 Nigeria, Tunisia, Gabon and Reunion patients samples as of yellow color of a tree node as it was illustrated on **figure 1A** phylogenetic tree.

Thus, it has theoretical to be the sample from 4 of them EPI_ISL_1169919, EPI_ISL_1170958, EPI_ISL_1170957 and EPI_ISL_1170956 categorized as clade of other (O) and it has not defined yet, whereas 7 of them were classified as GH clad distributions which include EPI_ISL_1899485,

EPI_ISL_1899485, EPI_ISL_1898814, EPI_ISL_1898554, EPI_ISL_1898258, EPI_ISL_1897648, EPI_ISL_1897646 and EPI_ISL_1897644, but the rest 3 samples were absolutely different and categorize as GR clad distributions which including EPI_ISL_1899281, EPI_ISL_1899047 and EPI_ISL_1897884 were confirmed as a new varaints of VUI202012/01/GRY (B.1.1.7). Pango linage B.1.1.7 was first detected in the UK and is now circulating in more than 134 countries, including Ethiopia.

We performed spike glycoprotein mutational ancestral relationship against to wild Wuhan genome and SARS-like-CoV. It has proven that the gradual mutational changes on spike protein in related its ancestors. Accordingly the GASAID, 2020, the first 4 genome samples of (EPI_ISL_1169919; EPI_ISL_1170956; EPI_ISL_1170957 and EPI_ISL_1170958 there was at about 30 amino acid changes along 97.9% identity against to bat/Yunnan/RaTG13/2013 whereas 232 aa changed with 81.36% identity in contradiction of SARS-like/Bat/Nanjing/SL-CoVZXC21/2015 and 226 aa changed with 81.86% against to SARS-like/Bat/Nanjing/SL-CoVZC45/2017 respectively, but only 4 aa was altered with 99.6 % identity of hCoV-19/Wuhan/WIV04/2019 on **Figure 1 (A1, B1, C1 and D1);** whereas the second round 10 genome samples of EPI_ISL_1899485; EPI_ISL_1899281; EPI_ISL_1899047; EPI_ISL_1898814; EPI_ISL_1898554; EPI_ISL_1898258; EPI_ISL_1897884; EPI_ISL_1897648; EPI_ISL_1897646; EPI_ISL_1897644 had about 33 amino acid changes along 97.4% identity against to bat/Yunnan/RaTG13/2013 whereas 234 aa changed with 81.2% identity in contradiction of SARS-like/Bat/Nanjing/SL-CoVZXC21/2015 and 228 aa changed with 81.7% against to SARS-like/Bat/Nanjing/SL-CoVZC45/2017 respectively on **Figure 1 (A2, B2, C2 and D2)** this implies that the mutational changes on structural spike was before a decades and continue to at present SARS-CoV-2. This finding showed that how the spike proteins changed over time as compared to the previous coronavirus and is probable that they will change rigorously.

SARS-CoV-2 Spike Dynamic Mutational Changes and its Antigenicity

The sum total of 14 SARS-CoV-2 genomes from Ethiopia was sequenced and deposited in GISAID datasets as of April/2021 [12, 13]. And a number of structural and none structural mutational changes were evaluated. In particular, the changes on Spike proteins that will have an effect on site of receptor binding and able to alter host receptors or antigenicity were considered for this analysis. Essentially, a 21 number of notable mutational variations were recorded that will have probability antigenicity effects and found on structural termini including G142A;Y144del(143); E180Q; E309Q; P330L; N440K; N501Y; A520S; A570D; D614G; P681H(674); T716I; A771S; S939F; S982A; T1006I; D1118H; K1073N; A1078S while none structural S13T(N-term); M1229I(C-term) as it had depicted on **Figure 2**. However, Spike-D614G, N501Y, Spike_M1229I, NSP12_P323L mutation were reflected as high prevalence of SARS-CoV-2 and circulating in Ethiopia. Some of the mutational changes that happened in Ethiopia exclusively, such as Spike-T478X, Spike-H69X, Spike-N440K and Spike-D614X.

Therefore, the genetic mutation changes are expected to vary depending on the immune response of the host cells and contumely fluctuations across each patient's reaction. Though most mutations could be

nonthreatening, the spike glycoprotein mutational changes of Spike_D614G help strengthen the viral survival capability [14]. The N501Y mutation might lead to the viral phenotypic character and its fusion or degradability of the receptor cells, whereas NSP5-S284G Non-structural protein-5 plays in the viral genome replication into the host for the formation of viral factories.

RBD of Spike Conserved Region and mAb Neutralizations Against Spike

It is important to investigate the function of the virus Receptor Binding Domain (RBD) and its motif and the targeted receptor host cells because the integration of molecular amino acid peptides and gene machinery gives us sufficient information about the mechanism of entry of virus and evasion mode of action into receptor cells and to know the mechanism of inhibitor drugs or to develop antibody neutralization and its vaccinations. So, searching for conserved domain of spike protein is one of the techniques to reveal the molecular evolution of genes and organisms. It has confirmed that the N-terminal domain (NTD) and C- terminal domain (CTD) of the S1 & S2 subunit of the Spike (S) proteins as reference QSM35600.1 accession number the sample was taken from one of Ethiopian patients were conserved from beta coronaviruses in the sarbecovirus subgenera (B lineage), and use the C-domain to bind their receptors with higher binding affinity than that of the previous SARS and MERS coronavirus and binding polypeptide sites of RBD and RBM has also noted. It enables itself to attach on the receptor cells perfectly [15]. Having to the conserved domain finding, S spike of SARS-CoV-2 of our sample (Ethiopia QSM35600.1) of fusion peptide site is highly conserved to pangolin coronavirus (QIQ54048.1), bat-SL-CoVZXC21, (AVP78042.1), HKU3 (Q3LZX1.1), SARS CoV-2 (PDB: 6VSB_A and 6ACC_A, 2021). However, exclusively SARS-CoV-2 has a functional polybasic (furin) cleavage site (R-arginine/S-threonine), which is absent in SARS coronavirus in **figure 3.** As a result, changes in the receptor domain affect its adhesiveness of affinity and the virus is more likely to be transmitted to humans easily as compared to the previous coronavirus. Annotated Spike protein (Ethiopia-QSM35588.1) structural modeling and ACE2 confirmation was done through modeler and taken its analogue crystallographic structure from the protein data bank (PDB: 6acc, EM 3.6 Angstrom) with RBD and Spike glycoprotein (PDB: 6acj, EM 4.2 Angstrom) in complex with host cell receptor ACE2. The interaction of the viral receptor binding domain and ACE-2 was revealed using molecular docking simulation through pyMol software. It has confirmed that the virus has four binding residues of Spike receptor binding sites, which are THR-486, TYR-436, ASN-473 and TYR-475 on **figure 3 B & D.** The interaction between Spike and ACE2 may be altered because of residual changes on the receptor binding sites and it could govern the responsibility of viral attachment and fusion activities [16].

Specifically, the mutational changes on PI_ISL_1898258 sample was detected as N440K amino acid changes and expected which will determine the neutralization of human antibody and vaccine escape mechanism because, as it was illustrated in **figure 3 (B1 & B2)** 7k8t PDB pyMol analysis ASN-440 glycoproteins of SARS-CoV 2 interact with SER-52 antibody might change the mode of action entirely and have antigenic-drift effects. This indicates how the viral infection could re-evasive the community through

go through the existence of antibodies or vaccines. Consequently, the total confirmed cases and number of deaths increased as per the national COVID-19 dash board as it has shown the graph in **figure 3 (F & G)**.

Conclusion And Discussion

The SARS-CoV-2 complete genome from COVID-19 positive patients of Ethiopia has ultimate evolutionary ancestral related to bat RaTG13 coronavirus followed by pangolin bat-SL-CoVZC45 and bat-SL-CoVZXC21, but genetic distance from MERS The clade distribution of the virus is supposed to be labeled as other GR, GH and O and its variants intended for new strains and VUI202012/01-GRY (B.1.1.7) Variants Interest was noted and Variants concerning VOC GH/501Y.V2 (B.1.351) of South Africa will be predicted. Hence we have recorded Spike N501Y mutation changes multiple times. Since the number of deaths, cases and hospitalization has been increased in Ethiopia, it is expected that Variant Concerns. The spike proteins of the SARS-CoV-2 had been conserved from its precursor of coronavirus and used the same binding sites against ACE-2. Mainly, D614G, D614X, N501Y, T478X, N440K and NSP5 S284G mutation had a significant impact in the country and caused high transmission rate. Thus, the existence of RT-PCR primer diagnosis kits and vaccines on the market may not be effective and could escape the immune system. The study highly recommended that the government responsible bodies should set strict preventative measurement to control the circulating variants in Ethiopia and further clinical and genomic study should be conducted.

Limitations

The study has limitations as;

- Minima genome sample sequences from Ethiopia during the time of this study.
- Selected 67 genomes were too small as compared to the world of more than one hundred thousand sequences.

Declarations

Ethics approval and consent to participate- **'Not applicable'**

List of abbreviation-

SARS-CoV-2	Sever Acute Respiratory Syndrome Coronavirus-2
ACE-2	Angiotensin Converting Enzyme-2
COVID-19	Corona Virus Infection Disease-19
INSA	Information Network Security Agency

MERS	Middle East Respiratory Syndrome
SARS	Sever Acute Respiratory Syndrome
MSA	Multiple Sequence Alignment
RBD	Receptor Binding Domain
mAbs	Monoclonal Antibodies
WHO	World Health Organization

Acknowledgement

Conflicts of Interest

It is declared that no conflict of interests among the authors

Data Availability Statement

We analyzed SARS-CoV-2 genome sequence data available from the NCBI-Nucleotide database https://www.ncbi.nlm.nih.gov/. The variants and genome sequence data and metadata of each COVID-19 positive patients are also taken from nextstrain datasets which are deposited at https://www.gisaid.org/. Annotated protein amino acid sequences are available at NCBI-Virus https://www.ncbi.nlm.nih.gov/labs/virus/vssi/#/. The analyzed Spike conserved region is found in

https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi. We Explored the 3D structure of proteins was retrieved from the PDB repository at https://www.rcsb.org/. COVID-19 number of cases and death records are available at https://covid19.who.int/region/afro/country/et.

References

1. World Health Organization. 2020a. Archived: WHO timeline—COVID-19. Available at https://www.who.int/news/item/27-04-2020-who-timeline—covid-19 (accessed 14 June 2020).

2. INSA. 2021. COVID net. INSA. https://www.covid19.et/covid-19/ (Accessed June 20/2021)

3. Kempen, J.H., Abashawl, A., Suga, H.K., Difabachew, M.N., Kempen, C.J., Debele, M.T., Menkir, A.A., Assefa, M.T., Asfaw, E.H., Habtegabriel, L.B. and Addisie, Y.S., 2020. SARS-CoV-2 serosurvey in Addis Ababa, Ethiopia. The American Journal of Tropical Medicine and Hygiene, 103(5), p.2022.

4. Korber, B., Fischer, W.M., Gnanakaran, S., Yoon, H., Theiler, J., Abfalterer, W., Hengartner, N., Giorgi, E.E., Bhattacharya, T., Foley, B. and Hastie, K.M., 2020. Tracking changes in SARS-CoV-2 spike: evidence that D614G increases infectivity of the COVID-19 virus. Cell, 182(4), pp.812-827.

5. Dejnirattisai, W., Zhou, D., Ginn, H.M., Duyvesteyn, H.M., Supasa, P., Case, J.B., Zhao, Y., Walter, T.S., Mentzer, A.J., Liu, C. and Wang, B., 2021. The antigenic anatomy of SARS-CoV-2 receptor binding domain. Cell, 184(8), pp.2183-2200.

6. Wang, Z., Schmidt, F., Weisblum, Y., Muecksch, F., Barnes, C.O., Finkin, S., Schaefer-Babajew, D., Cipolla, M., Gaebler, C., Lieberman, J.A. and Oliveira, T.Y., 2021. mRNA vaccine-elicited antibodies to SARS-CoV-2 and circulating variants. Nature, 592(7855), pp.616-622

7. Tamura, K., Stecher, G. and Kumar, S., 2021. MEGA11: Molecular Evolutionary Genetics Analysis version 11. Molecular biology and evolution, 38(7), pp.3022-3027.

8. Waterhouse, A.M., Procter, J.B., Martin, D.M., Clamp, M. and Barton, G.J., 2009. Jalview Version 2—a multiple sequence alignment editor and analysis workbench. Bioinformatics, 25(9), pp.1189-1191.

9. Schrödinger, L.L.C., 2015. The PyMOL molecular graphics system, version 1.8.

10. Cui, J., Li, F. and Shi, Z.L., 2019. Origin and evolution of pathogenic coronaviruses. Nature Reviews Microbiology, 17(3), pp.181-192.

11. Li, W., Shi, Z., Yu, M., Ren, W., Smith, C., Epstein, J.H., Wang, H., Crameri, G., Hu, Z., Zhang, H. and Zhang, J., 2005. Bats are natural reservoirs of SARS-like coronaviruses. Science, 310(5748), pp.676-679.

12. Raphael Tze Chuen Lee, Shi Shu Yuan, Ashar Malik, Frank Eisenhaber, GISAID Database Technical Group and Scientific Advisory Council

13. Elbe, S. and Buckland-Merrett, G., 2017. Data, disease and diplomacy: GISAID's innovative contribution to global health. Global challenges, 1(1), pp.33-46.

14. Korber, B., Fischer, W., Gnanakaran, S., Yoon, H., Theiler, J., Abfalterer, W., Hengartner, N., Giorgi, E., Bhattacharya, T., Foley, B., Hastie, K., Parker, M., Partridge, D., Evans, C., Freeman, T., de Silva, T., McDanal, C., Perez, L., Tang, H., Moon-Walker, A., Whelan, S., LaBranche, C., Saphire, E., Montefiori, D.,

Angyal, A., Brown, R., Carrilero, L., Green, L., Groves, D., Johnson, K., Keeley, A., Lindsey, B., Parsons, P., Raza, M., Rowland-Jones, S., Smith, N., Tucker, R., Wang, D. and Wyles, M., 2020. Tracking Changes in SARS-CoV-2 Spike: Evidence that D614G Increases Infectivity of the COVID-19 Virus. Cell, 182(4), pp.812-827.

15. Marchler-Bauer, A., Bo, Y., Han, L., He, J., Lanczycki, C.J., Lu, S., Chitsaz, F., Derbyshire, M.K., Geer, R.C., Gonzales, N.R. and Gwadz, M., 2017. CDD/SPARCLE: functional classification of proteins via subfamily domain architectures. Nucleic acids research, 45(D1), pp.D200-D203.

16. Ayele, A., Abdissa, B., Taye, D., Yemane, B. and Majumdar, R.S., 2020. COVID-2019 Genome Sequence Analysis: Phylogenetic Molecular Evolution and Docking of Structural Modelling of Receptor Binding Domain of S Protein in Active Site of ACE2. Computational Molecular Bioscience, 10(03), p.95.
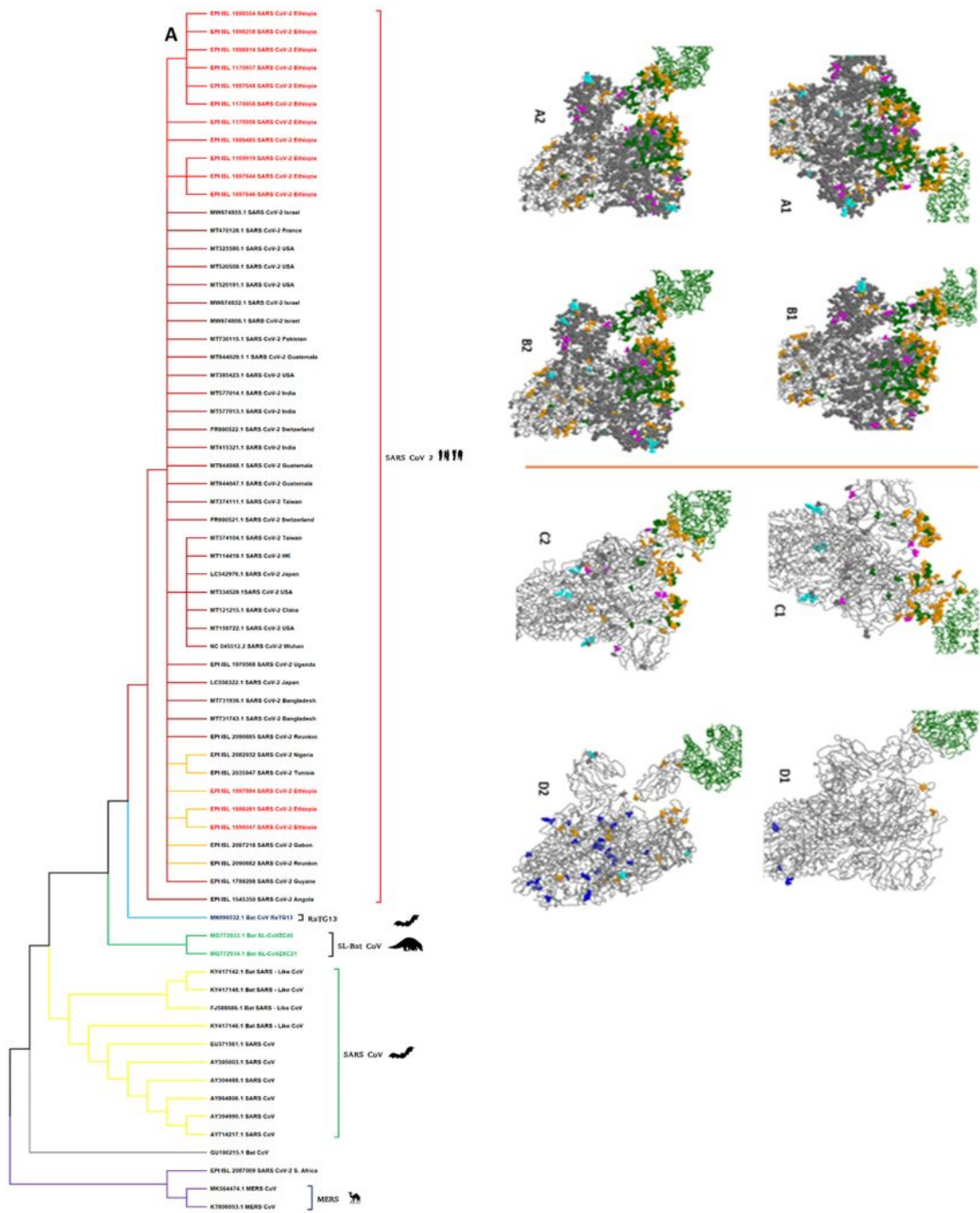
Figures

Figure 1

Pandemic SARS-CoV-2, Bat SARS-like coronavirus and MERS coronavirus A Phylogenetic tree was constructed using neighbor-joining statistical approaches and a single amino acid substitution method. The Poisson model along with phylogeny test of 1000 bootstrap replication methods (A). GISAID analysis of ACE-2 (deep green color) interaction with dynamical mutational amino acid changes in Ethiopia 14 samples which are colored in blue are common mutations that appeared more than 100 occurrences and

which colored in yellow have not occurred multiple times in amino acid sequence and are responsible for attachment, fusion, and antigenicity and neutralize the host cells. whereas the black one is not defined yet, insertion or deletion of amino acid residues are colored in cyan, removing a potential glycosylation site is colored magenta, on S of spike as query of Ethiopian COVID-19 positive patients genome sequence as reference of SAR S-like/Bat/Nanjing/SL-CoVZXC21/2015 (A1 & A2), SARS-like/Bat/Nanjing/SL-CoVZC45/2017 (B1 & B2) bat/Yunnan/RaTG13/2013 (C1 & C2) and hCoV-19/Wuhan/WIV04/2019 (D 1 & D2).
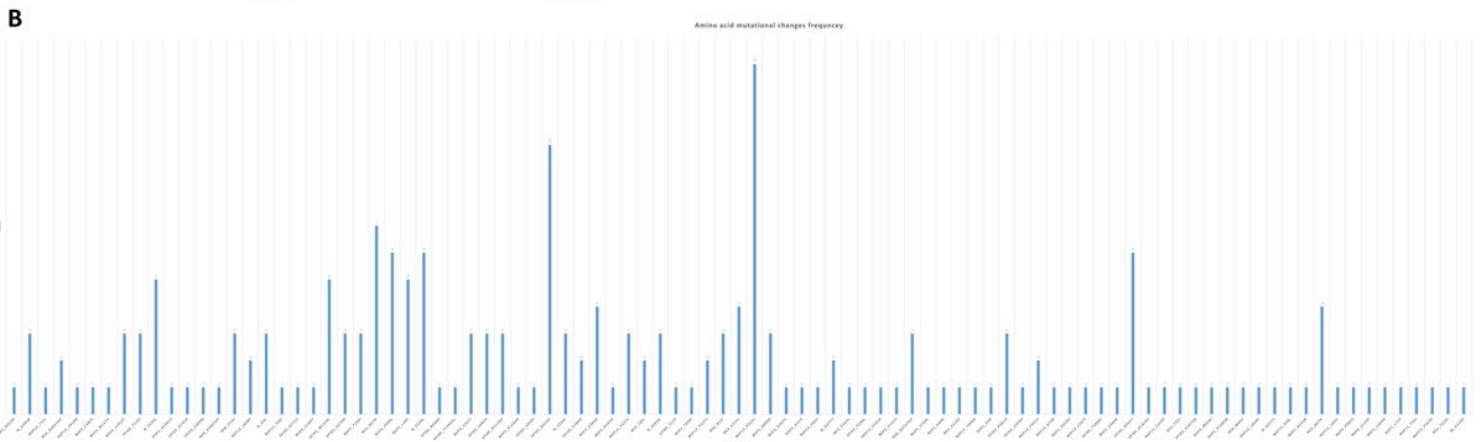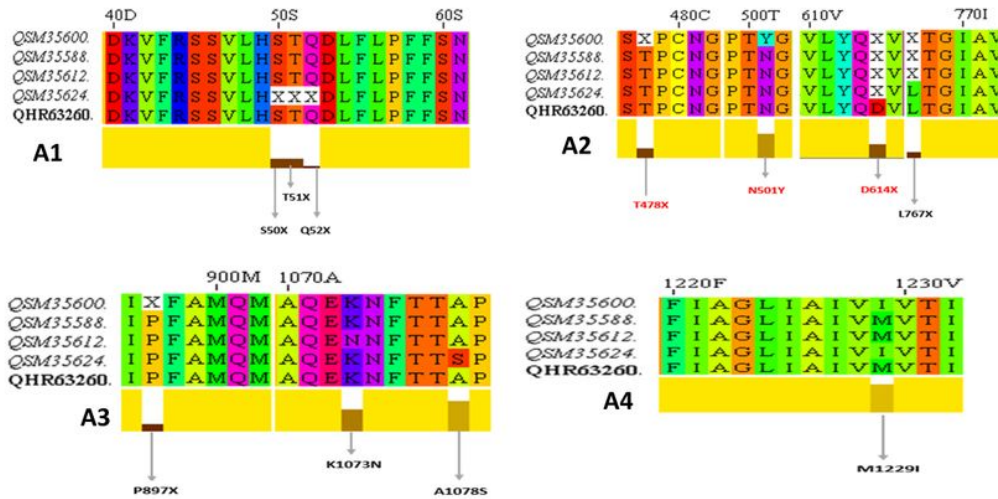

Figure 2

Multiple Sequences Alignment of SARS-CoV-2 Spike amino acid (aa) sequence of Ethiopia (QSM35588.1, QSM35600.1, QSM35612.1 and QSM35624.1) against wild SARS-CoV-2 Wuhan (QHR63260.2) More mutational changes that might have been responsible for the infectivity of the virus were colored in red (A1-A4). Total frequency of SARS-CoV-2 Spike amino acid mutational changes of COVID-19 positive sample in Ethiopia (B)
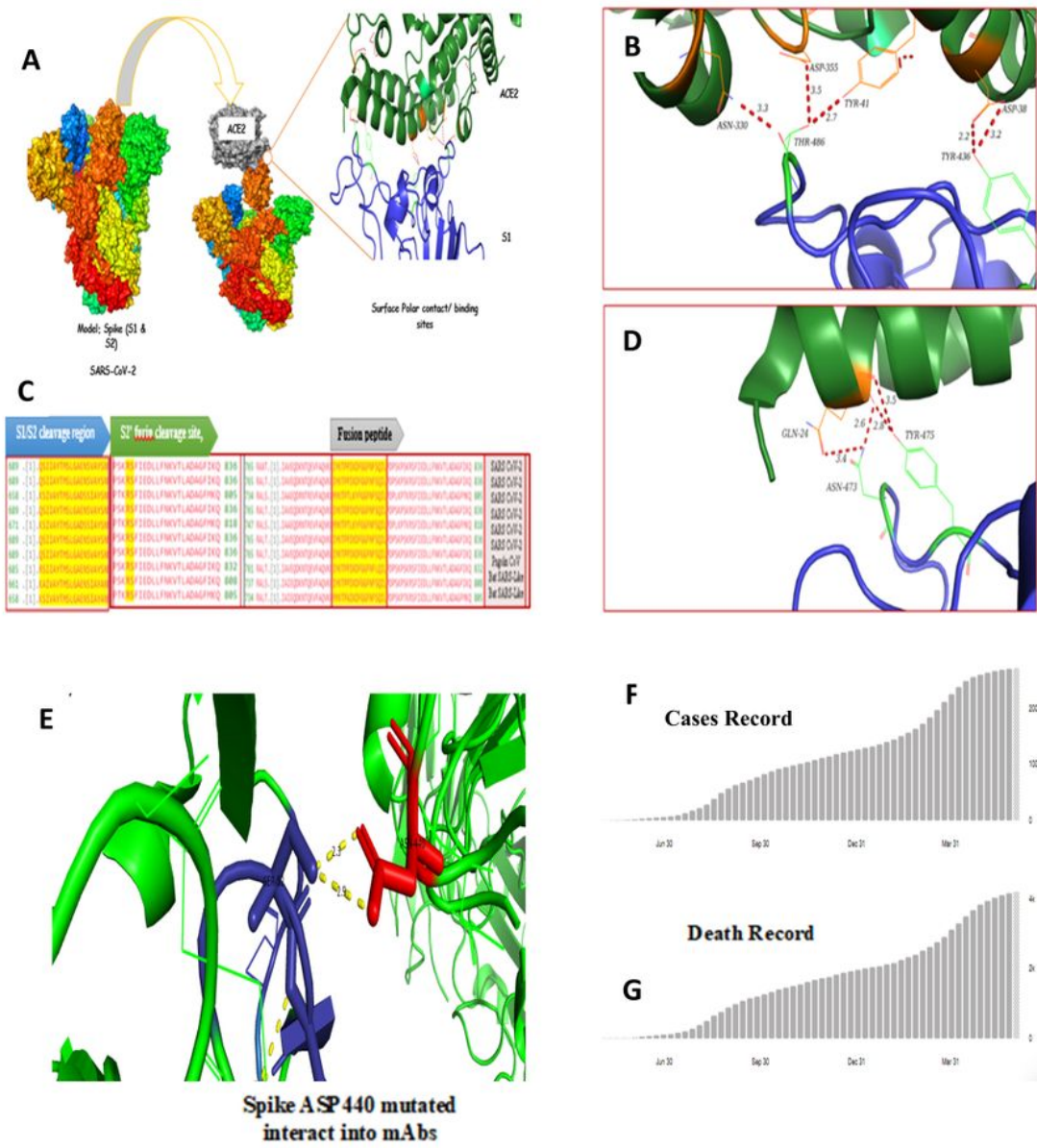
Figure 3

Spike glycoprotein (PDB: 6acj, EM 4.2 Angstrom) in complex with host cell receptor ACE2 (A) The amino acid residues and exact interaction between the Receptor Binding Domain of spike protein colored light blue and responsible residues for attachment and fusion THR-486, TYR-436, ASN-473 and TYR-475 and human Angiotensin converting Enzyme-2 as a host cell colored as deep green (B & D). Conserved Domain of Receptor Binding site of N-terminal domain (NTD) and C-terminal domain (CTD) of Spike glycoprotein

(QSM35600.1) where the S1/S2 cleavage region takes place as blue color, S2' furin cleavage site of membrane mediated fusion as indicated on green color and hydrophobic interaction of host cells as black color and all yellow highlighted as conserved region (C). The interaction between SARS CoV2 glycoprotein mutational changes at N440K red color (B2) with neutralizing antibody purple 7k8t PDB (E) The total number of COVID-19 cases to date (F) and total number of confirmed deaths (G) in Ethiopia as per COVID-19 dashboard

Supplementary Files

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

- TableS1.docx