

SARS-CoV-2 Genome Variants Epidemiology Surveillance in Ethiopia and Targeted Mutational Changes of Structural S Spike Proteins Through Computational Analysis

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

School of Basic Science and Research, Sharda University and Technology and Innovation Institute

Rita Majumdar

Sharda University School of Engineering and Technology

Yohannes Sitotaw

Ethiopian Biotechnology Institute

Tesfaye Adisu

Near East University: Yakin Dogu Universitesi

Research note

Keywords: Phylogenetic Tree, Conserved Domain, Clade, Variants Concern, Spike mutation Antigenicity

Posted Date: July 9th, 2021

DOI: <https://doi.org/10.21203/rs.3.rs-692993/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

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 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 concern VUI202012/01GRY B.1.1.7 which Pango lineage B.1.1.7 was originated from the UK. Despite 21 notable mutations, 71% D614G, 28% D614X, 35% N501Y and 21% NSP5 S284G mutation were occurred predominantly in our genome samples. and could be antigenicity and infectivity. Mutation on N440K was perceived 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 at around 4 million people out of the total half of billion 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 became 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 AstraZenec, Pfizer-Biotech vaccine [5]. Variants of B.1.351 N501Y also confirmed its resistance to therapeutically monoclonal antibodies [6]. Hence the country as Ethiopia is likely poor and very difficult to overcome those dynamic mutations without genomic surveillance checkpoints.

In this study, we constructed evolutionary relatedness of 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 was 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 positive were retrieved from GISAID datasets and performed MSA through blastn techniques 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 non structural protein mutational changes and other epidemiology analysis was done through GISAID CoVsurver platforms <https://www.gisaid.org/>. Protein amino acid sequence was analyzed using Jalview (version 2.10.0.) [8] and conserved domain (CD) analysis platforms. Finally, interaction between Spike receptor binding sites [PDB: 6acj] and ACE-2 as well between mutated sites [PDB: 7k8t] and mAbs were analyzed through 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 phylogenetic 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 analysis on Fig. 1A. This finding provisionally took a scientific evidence as per [10, 11], reported *Rhinolophus affinis* was considered a natural reservoir of the coronavirus and supposed to a source of future epidemic.

Even though 11 patients SARS CoV-2 genome samples from Ethiopia were grouped as in same genetic distance as it marked in red color of tree, it is reflected that sub family member of EPI ISL 1897884, EPI ISL 1899281 and EPI ISL 1899047 were clustered with same sub-group and genotype association to SARS CoV-2 Nigeria, Tunisia, Gabon and Reunion patients samples as of yellow color of a tree node as it is illustrated on Fig. 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 the next 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 3 samples were absolutely different and categorize as GR clad distributions which consist of EPI_ISL_1899281, EPI_ISL_1899047 and EPI_ISL_1897884 were

confirmed as a new strains of VUI202012/01 GRY (B.1.1.7). Pango lineage B.1.1.7 was first detected in UK and now circulating more than 134 countries including Ethiopia.

We had also performed spike glycoprotein mutational ancestral relationship against to Wuhan genome and SARS. It has proven that the gradual mutational changes on spike protein in related its ancestors. Accordingly the GISAID, 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 Fig. 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 Fig. 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 shown that how the spike proteins changed over the time as compared to the previous coronavirus and probable that will change rigorously.

SARS CoV 2 Spike dynamic mutational changes and its antigenicity

The sum total of 14 SARS CoV-2 genome 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. Especially, the changes on S gene or Spike proteins those will have effect on site of receptor binding and able to alter host receptors or antigenicity was considered for this analysis. Essentially 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 and none structural S13T(N-term) M1229I(C-term) as it has depicted on Fig. 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 mutational changes those happened in Ethiopia exclusively such as Spike T478X, Spike H69X, Spike, N440K and Spike D614X.

Therefore, the genetic mutation changes is expected vary depend open the immune response of the host cells and contumely fluctuations across each patients reaction. Though most mutational could be as nonthreatening, the spike glycoproteins mutational changes of Spike_D614G help strengthen the viral survival capability [14]. N501Y mutation might be lead 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 to 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 give us sufficient of 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 that searching of conserved domain of spike protein is one of the technique 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 were 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 enable 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 on Fig. 3. As a result of changes in the receptor domain affects its adhesiveness of affinity and the virus is more likely to be transmitted to humans easy as compare to the previous coronavirus. Annotated Spike protein (Ethiopia-QSM35588.1) structural modeling and ACE2 confirmation was done through modeler and taken its analogues crystallographic structure from 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 viral receptor binding domain and i-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 Fig. 3B & 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 samples was detected as N440K amino acid changes and expected that which will determine the notarization of human antibody and vaccine escape mechanism because as it was illustrated on Fig. 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. This indicates how the viral infection could re-evasive the community through go through the existence antibodies or vaccines. Consequently, the total confirmed cases and number of death increase as per the national COVID - 19 dash board as it has shown the graph on Fig. 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 MERS coronavirus. The clade distribution of the virus supposed to be labeled as other GR, GH and O and its variants intended to new strains and VUI202012/01 GRY (B.1.1.7) variants concern was noted an predictable that VOC GH/501Y.V2 (B.1.351) hence we have recorded Spike N501Y mutation multiple times. Since the number of death, 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 same binding sites against to ACE-2. Mainly, D614G, D614X, N501Y, T478X, N440K and NSP5 S284G mutation had significance impact in the country. Thus, an existence RT-PCR primer diagnosis kits and vaccine on the market may not effective and could escape 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 conduct.

Limitations

The study has limitation as;

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

List Of Abbreviations

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

MSA Multiple Sequence Alignment

RBD Receptor Binding Domain

WHO World Health Organization

Declarations

Ethics approval and consent to participate

'Not applicable'

Funding

'Not applicable'

Consent for publication

'Not applicable'

Author's contributions

(Abaysew Ayele: proposing idea, data analysis using software and write up the manuscript, Rita Majumdar: supervise and advise, Yohannes Stitotaw: retrieved data, Tesfaye Adisu: literature review)

Acknowledgement

We would like thank the developer and curator in NCBI and GISAID dataset and virtual big data analytic platforms. Then we have special thanks for TECHIN who help and support us to do this SARS CoV-2 genomic research. Finally, we acknowledgement for those who have safe and deposit our analytic data into their own servers for fixed time and who gave us their unreserved scientific comments.

Conflicts of Interest

It is declared that no conflict of interests among the authors

Data Availability Statement

That we analyzed SARS CoV-2 genome sequence data available from NCBI- Nucleotide database <https://www.ncbi.nlm.nih.gov/>. The variants and genome sequence data and metadata of each COVID-19 positive patients also taken from nextstrain datasets which is deposited at <https://www.gisaid.org/>. Annotated protein amino acid sequences available at NCBI-Virus <https://www.ncbi.nlm.nih.gov/labs/virus/vssi/#/> and 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 PDB repository in <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 Jan 20/2021)
3. Kempen, J. H., Abashawl, A., Suga, H. K., Nigussie Difabachew, M., Kempen, C. J., Tesfaye Debele, M., Menkir, A. A., Assefa, M. T., Asfaw, E. H., Habtegabriel, L. B., Sitotaw Addisie, Y., Nilles, E. J., & Longenecker, J. C. (2020). SARS-CoV-2 Serosurvey in Addis Ababa, Ethiopia. *The American journal of tropical medicine and hygiene*, 103(5), 2022–2023. <https://doi.org/10.4269/ajtmh.20-0816>
4. Korber, B., Fischer, W., Gnanakaran, S.G., Yoon, H., Theiler, J., Abfalterer, W., Foley, B., Giorgi, E.E., Bhattacharya, T., Parker, M.D. and Partridge, D.G., 2020. Spike mutation pipeline reveals the emergence of a more transmissible form of SARS-CoV-2. *BioRxiv*.
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. Kumar S, Stecher G, Li M, Nnyaz C, Tamura K. MEGA X: Molecular Evolutionary Genetics Analysis across Computing Platforms. *Mol Biol Evol*. 2018 Jun 1;35(6):1547-1549. doi: 10.1093/molbev/msy096. PMID: 29722887; PMCID: PMC5967553.
8. Waterhouse, A.M., Procter, J.B., Martin, D.M.A., Clamp, M. and Barton, G.J. (2009) Jalview Version 2: A Multiple Sequence Alignment and Analysis Workbench. *Bioinformatics*, 25, 1189-1191. <https://doi.org/10.1093/bioinformatics/btp033>
9. Schrodinger, LLC (2015) The PyMOL Molecular Graphics System, Version 2.3.0.
10. Cui, J., Li, F. and Shi, Z.-L. (2019) Origin and Evolution of Pathogenic Coronaviruses. *Nature Reviews Microbiology*, 17, 181-192. <https://doi.org/10.1038/s41579-018-0118-9>
11. Li, W., Shi, Z., Yu, M., Ren, W., Smith, C., Epstein, J.H., Wang, H., Crameri, G., Hu, Z. and Zhang, H. (2005) Bats Are Natural Reservoirs of SARS-Like Coronaviruses. *Science*, 10, 676-679. <https://doi.org/10.1126/science.1118391>
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*, 1:33-46. DOI:10.1002/gch2.1018 PMCID: 31565258
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.e19.

15. Marchler-Bauer, A., Bo, Y., Han, L., He, J., Lanczycki, C., Lu, S., Chitsaz, F., Derbyshire, M., Geer, R., Gonzales, N., Gwadz, M., Hurwitz, D., Lu, F., Marchler, G., Song, J., Thanki, N., Wang, Z., Yamashita, R., Zhang, D., Zheng, C., Geer, L. and Bryant, S., 2016. 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. (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, 95-110. doi: 10.4236/cmb.2020.103007.

Figures

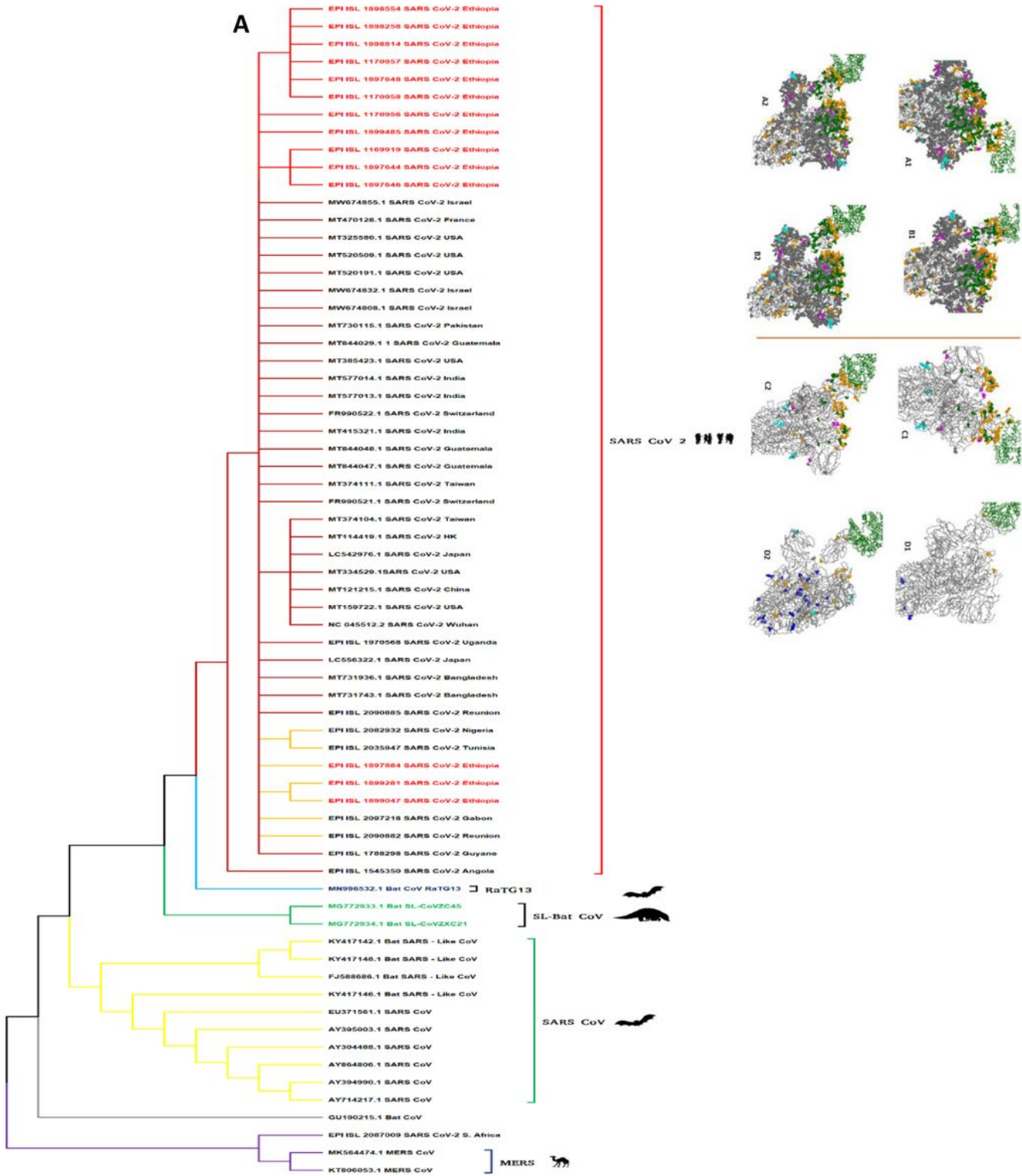


Figure 1

Pandemic SARS CoV-2, Bat SARS-like coronavirus and MERS coronavirus Phylogenetic tree was constructed using neighbor-joining statistical approaches and a single amino acid substitution methods. Poisson model along 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 colored in blue are common mutations that appeared more than 100 occurrence and which colored

in yellow are not occurred multiple times in amino acid sequence and responsible for attachment, fusion, and antigenicity and neutralize the host cells. whereas the black one not defined yet, insertion or deletion of amino acid residues are colored in cyan, remove a potential glycosylation site are 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 Sequence Alignment of SARS CoV-2 Spike amino acid sequence of Ethiopia (QSM35588.1, QSM35600.1, QSM35612.1 and QSM35624.1) against to itself SARS CoV-2 Wuhan (QHR63260.2) More mutational changes that might have responsible on the infectivity of the virus were colored in red (A1-A4). Frequency of SARS CoV 2 Spike Amino acid mutational changes of positive sample in Ethiopia (B)

(QSM35600.1) where the S1/S2 cleavage region is 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).and the interaction between SARS CoV2 glycoprotein mutational changes at N440K red color (B2) with neutralizing antibody purple 7k8t PDB (E) Total number of COVID-19 cases to date (F) and total number of confirmed death (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](#)