PLOS ONE

RESEARCH ARTICLE

Most frequent South Asian haplotypes of *ACE2* share identity by descent with East Eurasian populations

Anshika Srivastava[1], Rudra Kumar Pandey[1], Prajjval Pratap Singh[1], Pramod Kumar[2], Avinash Arvind Rasalkar[3], Rakesh Tamang[4], George van Driem[5], Pankaj Shrivastava[6], Gyaneshwer Chaubey[1]*

1 Department of Zoology, Cytogenetics Laboratory, Banaras Hindu University, Varanasi, India, **2** National Centre for Disease Control, Delhi, India, **3** Redcliffe Life Sciences Pvt Ltd, Electronic City, Noida, Uttar Pradesh, India, **4** Department of Zoology, University of Calcutta, Kolkata, India, **5** Institut für Sprachwissenschaft, Universität Bern, Bern, Switzerland, **6** Department of Home (Police), DNA Fingerprinting Unit, State Forensic Science Laboratory, Government of MP, Sagar, India

* gyaneshwer.chaubey@bhu.ac.in

OPEN ACCESS

Citation: Srivastava A, Pandey RK, Singh PP, Kumar P, Rasalkar AA, Tamang R, et al. (2020) Most frequent South Asian haplotypes of *ACE2* share identity by descent with East Eurasian populations. PLoS ONE 15(9): e0238255. https://doi.org/10.1371/journal.pone.0238255

Editor: Alessandro Achilli, Universita degli Studi di Pavia, ITALY

Received: April 27, 2020

Accepted: August 12, 2020

Published: September 16, 2020

Peer Review History: PLOS recognizes the benefits of transparency in the peer review process; therefore, we enable the publication of all of the content of peer review and author responses alongside final, published articles. The editorial history of this article is available here: https://doi.org/10.1371/journal.pone.0238255

Data Availability Statement: The used data is available at https://evolbio.ut.ee/CGgenomes.html.

Funding: This work is supported by the National Geographic Explorer grant HJ3-182R-18. Redcliffe

Abstract

It was shown that the human Angiotensin-converting enzyme 2 (ACE2) is the receptor of recent coronavirus SARS-CoV-2, and variation in this gene may affect the susceptibility of a population. Therefore, we have analysed the sequence data of *ACE2* among 393 samples worldwide, focusing on South Asia. Genetically, South Asians are more related to West Eurasian populations rather than to East Eurasians. In the present analyses of *ACE2*, we observed that the majority of South Asian haplotypes are closer to East Eurasians rather than to West Eurasians. The phylogenetic analysis suggested that the South Asian haplotypes shared with East Eurasians involved two unique event polymorphisms (rs4646120 and rs2285666). In contrast with the European/American populations, both of the SNPs have largely similar frequencies for East Eurasians and South Asians, Therefore, it is likely that among the South Asians, host susceptibility to the novel coronavirus SARS-CoV-2 will be more similar to that of East Eurasians rather than to that of Europeans.

Introduction

The novel coronavirus SARS-CoV-2, the causative agent of the ongoing pandemic of COVID-19, today presents one of the major challenges to humanity [1]. Recent studies have effectively demonstrated that the Angiotensin-converting enzyme 2 (ACE2) encoded by a gene located on the X-chromosome is the host receptor for the virus [1, 2]. A decreased level of *ACE2* expression mitigates the severity of the disease. The over-expression or a unique genetic polymorphism of the receptor among Asians have been ruled out in a recent study [3, 4]. ACE2 also maintains cardiovascular homeostasis and electrolyte balance and protects against lung injury by acid aspiration [5]. A comprehensive understanding of *ACE2* variations among various ethnic groups has hitherto been largely unknown.

Life Sciences Pvt Ltd. India provided support in the form of salaries for author AR. The specific roles of this author is articulated in the 'author contributions' section. The funders had no role in study design, data collection and analysis, decision to publish or preparation of the manuscript.

Competing interests: One of our author AR is full time employee of Redcliffe Life Sciences Pvt Ltd. India. This does not alter our adherence to PLOS ONE policies on sharing data and materials.

The South Asia subcontinent harbours diverse and endogamous ethnic groups [6]. Most of the genomes of South Asia are autochthonous but show a considerable amount of sharing with East and West Eurasia [7]. However, when we compare overall genome sharing with East *vs*. West Eurasia, South Asians show greater genetic affinity with West Eurasia [8–10]. The only exception is Tibeto-Burman speaking populations, who share a large amount of ancestry with East Eurasia [11]. The genetic structure of *ACE2* haplotypes among South Asian populations is not known. Therefore, we have analysed the whole genome data of South Asians with respect to various world populations for *ACE2* published elsewhere [12, 13] (S1 Table).

Materials and methods

The research has been approved by the Institutional Ethical Committee of Banaras Hindu University, Varanasi, India. To analyse the *ACE2* among various populations, we have extracted the sequences from the published datasets [12, 13], by using PLINK 1.9 [14]. It has been shown that the 1000 genome dataset for South Asia does not capture the complete South Asian variation, mainly due to unsampled Austroasiatic populations [15]. Hence, we analysed Pagani et al. [12] by way of primary data and further confirmed the results with the 1000 genome data [13]. We extracted 447 samples designated as a diversity set panel in the Pagani et al. data [12]. After excluding samples from Africa, Sahul and relatives up to the second degree, we used 393 samples in all our analyses (S1 Table). A total of 248 polymorphisms were observed in the Pagani et al. data [12] (S2 Table). LD maps for each of the groups were analysed from Haploview [16] (S1 Fig). For both of the datasets, we converted plink file to fasta file (ped to IUPAC) from customised script. Phasing of the data, the calculation of population-wise genetic distances, and Arlequin and Network input files were generated by DnaSP v 6 [17]. The neighbour joining (NJ) tree was constructed by MEGA-X [18] (Fig 1A). Nei's genetic distances and pairwise differences were calculated from Arlequin 3.5 [19] and plotted by R v 3.1 [20] (Fig 1B and S2 Fig). Network v5 [21] and Network publisher were used to construct the median joining (MJ) networks (Fig 2 and S3 Fig). The spatial map of rs4646120 and rs2285666 were drawn from *PGG* toolkit (S4 Fig) [22].

Result and discussion

Our pooled data have yielded 248 high quality polymorphisms (S2 Table). In the LD (linkage disequilibrium) plot analysis, significant LD blocks of different sizes were present among Caucasus, Central Asians, South Asians, mainland Southeast Asians, insular Southeast Asians and Siberians (S1 Fig). Europeans showed the lowest level of LD. We have used a haplotype based approach for the comparison. In contrast with the genome-wide analysis [8–10], the NJ (Neighbour Joining) tree based on *F*st distances clustered South Asians together with insular and mainland Southeast Asian populations (Fig 1A). This unexpected result suggested closer a genetic affinity of South Asians with East Eurasians for *ACE2*. The pairwise difference analysis suggested lower diversity for South Asian, Southeast Asian and Siberian populations (Fig 1B). Similarly, the 1000 genome populations showed the lowest diversity for East Asian populations (S2 Fig).

The phylogenetic analysis of various haplotypes among studied populations helped to identify the SNPs responsible for the affinity of South Asians with East Eurasians (Fig 2 and S3 Fig). Three major distinct haplotypes were observed. Haplotype 1 (ht1) was more common in West Eurasians, including Central Asian populations, whereas haplotype 2 (ht2) was frequent among East Eurasians, South Asians and Americans (Fig 2 and S3 Fig). Haplotype 3 (ht3) was harboured mainly by East Eurasians and South Asians. The haplotype 2 (ht2) originated from SNP rs4646120, whereas ht3 was derived from SNP rs2285666. Phylogenetically both of these

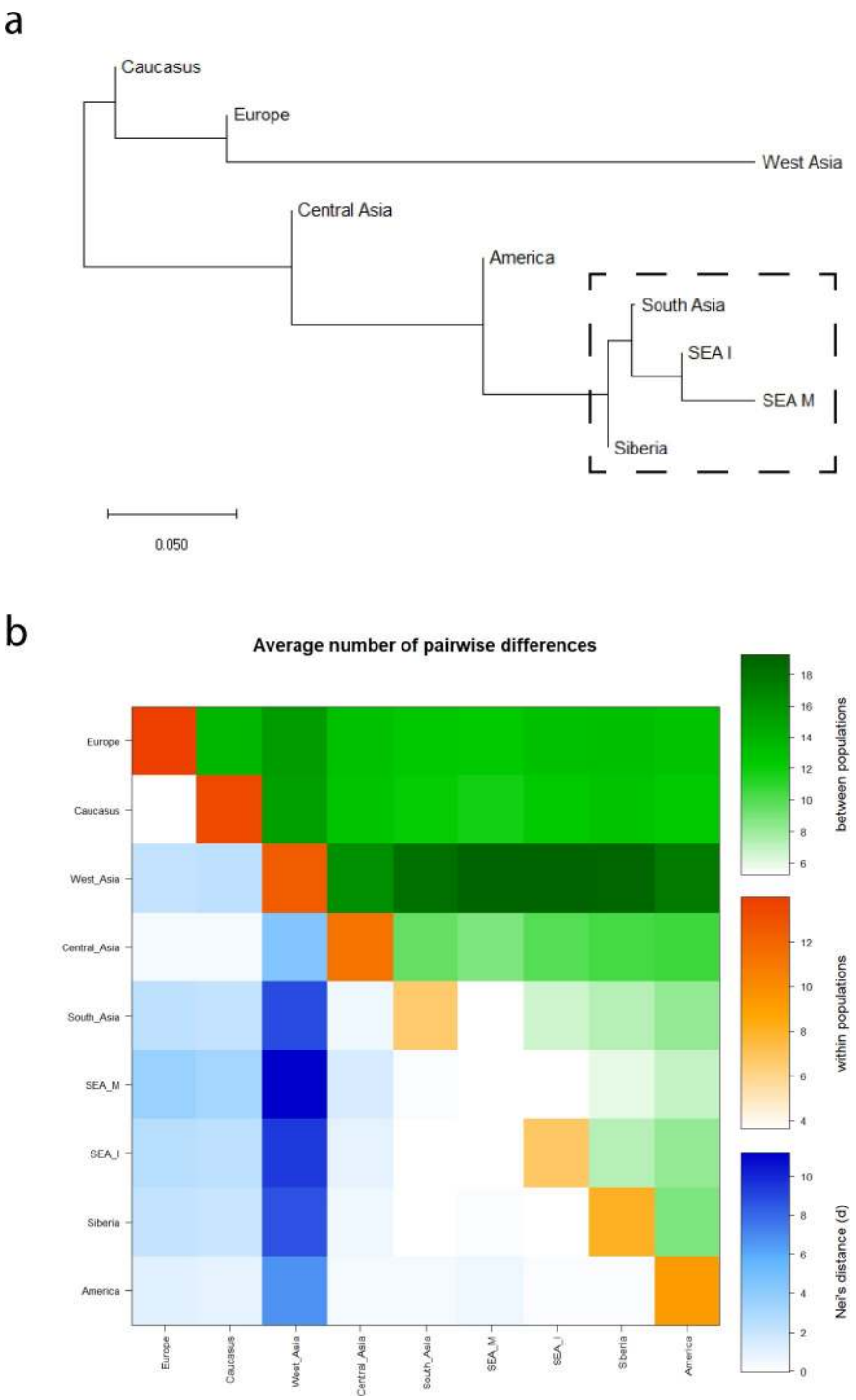

Fig 1. a) The Neighbour-Joining (NJ) tree showing the genetic relationship of the studied populations. The figure was drawn from the *F*st distances obtained from the haplotype analysis of Eurasian populations. b) Heat map showing the intra- and inter-population variation measured by average pairwise sequence differences of the ACE2 gene. The average pairwise differences between populations are shown in the upper triangle of the matrix (green). The average number of pairwise differences within each population group are shown along the diagonal (orange). The differences between populations based on Nei's genetic distances are depicted in lower triangle of the matrix (blue). The obtained values of various parameters have been shown at the color scales.

https://doi.org/10.1371/journal.pone.0238255.g001

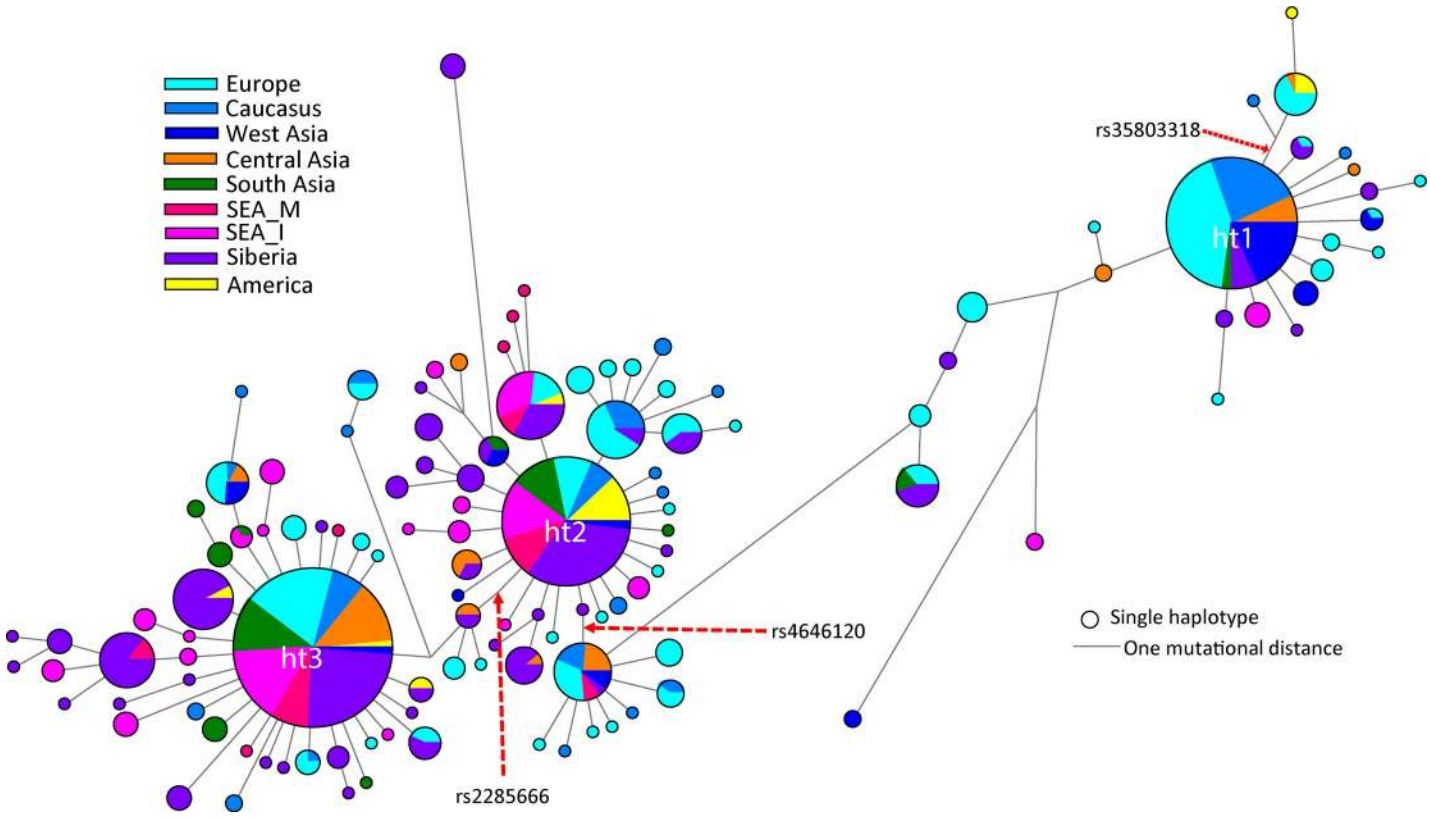

Fig 2. The median joining (MJ) network of 142 haplotypes belonging to gene ACE2. Circle sizes are proportional to the number of samples with that haplotype. The three most common haplotypes are marked. The three important SNPs studied in details have been marked in the figure. We used median joining method implemented in the NETWORK programme ver. 5.

https://doi.org/10.1371/journal.pone.0238255.g002

SNPs play a key role in the distinction between East and West Eurasian populations (Fig 2 and S3 and S4 Figs). Interestingly, the most frequent haplotypes of South Asia involve these SNPs. A recent study has also highlighted the highest frequency of this SNP (rs 2285666) among Chinese populations (0.5) as well as significant frequency differences among 1000 genome populations (S4 Fig) [4]. In our study, we also found high frequency (0.6) of this SNP among South Asians (S2 Table and S4 Fig). Moreover, we also found that a synonymous coding region variant rs35803318 was most frequent among Americans (0.15), followed by Europeans (0.055), Caucasians (0.051) and Central Asians (0.021), whilst this site was not polymorphic for West Asians, South Asians, Southeast Asians and Siberians (S2 Table).

Phylogenetic analysis has suggested that the majority of South Asian samples share with East Eurasians the monophyletic haplotypes 2 and 3 by the unique polymorphism events (rs4646120) and (rs2285666). Recent studies have suggested that the reference allele has a reduced *ACE2* expression of up to 50%, resulting in greater severity of a SARS-CoV-2 infection [23–25]. Additionally, a synonymous coding region variant rs35803318 was also significantly more polymorphic among Americans and Europeans than among South Asians. Hence, it is likely that among South Asians, the host susceptibility to the novel coronavirus SARS-CoV-2 more closely resembles that of East/Southeast Asians rather than that of Europeans or Americans.

Supporting information

S1 Fig. The LD (linkage disequilibrium) plots of ACE2 gene of various studied populations. Shading from white to red indicates the intensity of r^2 from 0 to 1. Strong LD is represented by a high percentage (>80) and a darker red square.
(TIF)

S2 Fig. Heat map showing the intra- and inter-population variation measured by average pairwise sequence differences of the ACE2 gene among 1000 genome populations. The average pairwise differences between populations are shown in the upper triangle of the matrix (green). The average number of pairwise differences within each population group are shown along the diagonal (orange). The differences between populations based on Nei's genetic distances are depicted in lower triangle of the matrix (blue). The populations are grouped in to superpopulations e.g. European, South Asian, East Asian and American.
(TIF)

S3 Fig. The median joining (MJ) network of 491 haplotypes belonging to gene ACE2 among 1000 genome populations. Circle sizes are proportional to the number of samples with that haplotype. The three most common haplotypes are marked. All three SNPs studied in detail have been marked by arrow. We used median joining method implemented in the NETWORK programme ver. 5.
(TIF)

S4 Fig. The spatial distribution of alleles of rs4646120 and rs2285666 among 1000 genome populations. The map was obtained from the *PGG* toolkit implemented in the https://www.pggsnv.org/.
(TIF)

S1 Table. The number of samples from each of the region used in the analysis. The number of South Asian groups shown with their linguistic affiliations.
(PDF)

S2 Table. The details of 248 polymorphic loci extracted from analysed data. The frequencies of alternate alleles for each loci and group have been mentioned in the table. The three important SNPs studied in details have been highlighted.
(XLSX)

Acknowledgments

We thank to both of the reviewers and the Editor for their constructive suggestions.

Author Contributions

Conceptualization: Rakesh Tamang, Gyaneshwer Chaubey.

Data curation: Rudra Kumar Pandey, Prajjval Pratap Singh, Pramod Kumar, Avinash Arvind Rasalkar, Gyaneshwer Chaubey.

Formal analysis: Anshika Srivastava, Rudra Kumar Pandey, Pankaj Shrivastava, Gyaneshwer Chaubey.

Investigation: Pramod Kumar, Rakesh Tamang, Gyaneshwer Chaubey.

Project administration: Gyaneshwer Chaubey.

Supervision: Gyaneshwer Chaubey.

Validation: Rudra Kumar Pandey, Prajjval Pratap Singh.

Writing – original draft: Anshika Srivastava, Gyaneshwer Chaubey.

Writing – review & editing: Prajjval Pratap Singh, Pramod Kumar, George van Driem.

References

1. Lu R, Zhao X, Li J, Niu P, Yang B, Wu H, et al. Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. The Lancet. 2020; 395:565–74.
2. Zhou P, Yang X-L, Wang X-G, Hu B, Zhang L, Zhang W, et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature. 2020; 579:270–3. https://doi.org/10.1038/s41586-020-2012-7 PMID: 32015507
3. Chen Y, Shan K, Qian W, Qian W. Asians do not exhibit elevated expression or unique genetic polymorphisms for ACE2, the cell-entry receptor of SARS-CoV-2. Preprints 2020. 2020020258 https://doi.org/10.20944/preprints202002.0258.v2
4. Cao Y, Li L, Feng Z, Wan S, Huang P, Sun X, et al. Comparative genetic analysis of the novel coronavirus (2019-nCoV/SARS-CoV-2) receptor ACE2 in different populations. Cell Discov. 2020; 6:1–4. https://doi.org/10.1038/s41421-019-0132-8
5. Yang J, Zheng Y, Gou X, Pu K, Chen Z, Guo Q, et al. Prevalence of comorbidities in the novel Wuhan coronavirus (COVID-19) infection: a systematic review and meta-analysis. Int J Infect Dis. 2020.
6. Nakatsuka N, Moorjani P, Rai N, Sarkar B, Tandon A, Patterson N, et al. The promise of discovering population-specific disease-associated genes in South Asia. Nat Genet. 2017; 49:1403. https://doi.org/10.1038/ng.3917 PMID: 28714977
7. Chaubey G, Metspalu M, Kivisild T, Villems R. Peopling of South Asia: investigating the caste-tribe continuum in India. BioEssays News Rev Mol Cell Dev Biol. 2007; 29:91–100.
8. Xing J, Watkins WS, Hu Y, Huff CD, Sabo A, Muzny DM, et al. Genetic diversity in India and the inference of Eurasian population expansion. Genome Biol. 2010; 11:R113. https://doi.org/10.1186/gb-2010-11-11-r113 PMID: 21106085
9. Metspalu M, Romero IG, Yunusbayev B, Chaubey G, Mallick CB, Hudjashov G, et al. Shared and unique components of human population structure and genome-wide signals of positive selection in South Asia. Am J Hum Genet. 2011; 89:731–44. https://doi.org/10.1016/j.ajhg.2011.11.010 PMID: 22152676
10. Reich D, Thangaraj K, Patterson N, Price AL, Singh L. Reconstructing Indian population history. Nature. 2009; 461:489–94. https://doi.org/10.1038/nature08365 PMID: 19779445
11. Chaubey G, Metspalu M, Choi Y, Mägi R, Romero IG, Soares P, et al. Population Genetic Structure in Indian Austroasiatic speakers: The Role of Landscape Barriers and Sex-specific Admixture. Mol Biol Evol. 2011; 28:1013–24. https://doi.org/10.1093/molbev/msq288 PMID: 20978040
12. Pagani L, Lawson DJ, Jagoda E, Mörseburg A, Eriksson A, Mitt M. Genomic analyses inform on migration events during the peopling of Eurasia. Nature. 2016;538. https://doi.org/10.1038/nature19792 PMID: 27654910
13. 1000 Genomes Project Consortium, Durbin RM, Abecasis GR, Altshuler DL, Auton A, Brooks LD, et al. A map of human genome variation from population-scale sequencing. Nature. 2010; 467:1061–73. https://doi.org/10.1038/nature09534 PMID: 20981092
14. Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ. Second-generation PLINK: rising to the challenge of larger and richer datasets. GigaScience. 2015; 4:1–16. https://doi.org/10.1186/2047-217X-4-1
15. Sengupta D, Choudhury A, Basu A, Ramsay M. Population Stratification and Underrepresentation of Indian Subcontinent Genetic Diversity in the 1000 Genomes Project Dataset. Genome Biol Evol. 2016; 8:3460–70. https://doi.org/10.1093/gbe/evw244 PMID: 27797945
16. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD and haplotype maps. Bioinforma Oxf Engl. 2005; 21:263–5.
17. Rozas J, Ferrer-Mata A, Sánchez-DelBarrio JC, Guirao-Rico S, Librado P, Ramos-Onsins SE, et al. DnaSP 6: DNA sequence polymorphism analysis of large data sets. Mol Biol Evol. 2017; 34:3299–302. https://doi.org/10.1093/molbev/msx248 PMID: 29029172
18. Kumar S, Stecher G, Li M, Knyaz C, Tamura K. MEGA X: molecular evolutionary genetics analysis across computing platforms. Mol Biol Evol. 2018; 35:1547–9. https://doi.org/10.1093/molbev/msy096 PMID: 29722887

19. Excoffier L, Lischer HEL. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. Mol Ecol Resour. 2010; 10:564–7. https://doi.org/10.1111/j.1755-0998.2010.02847.x PMID: 21565059

20. R-Core-Team. R: A language and environment for statistical computing. Vienna: R Foundation for Statistical Computing; 2012. http://www.R-project.org/.

21. Bandelt H-J, Forster P, Röhl A. Median-joining networks for inferring intraspecific phylogenies. Mol Biol Evol. 1999; 16:37–48. https://doi.org/10.1093/oxfordjournals.molbev.a026036 PMID: 10331250

22. Zhang C, Gao Y, Ning Z, Lu Y, Zhang X, Liu J, et al. PGG.SNV: understanding the evolutionary and medical implications of human single nucleotide variations in diverse populations. Genome Biol. 2019; 20:215. https://doi.org/10.1186/s13059-019-1838-5 PMID: 31640808

23. Asselta R, Paraboschi EM, Mantovani A, Duga S. ACE2 and TMPRSS2 variants and expression as candidates to sex and country differences in COVID-19 severity in Italy. 2020.

24. Wu Y, Li J, Wang C, Zhang L, Qiao H. The ACE 2 G8790A Polymorphism: Involvement in Type 2 Diabetes Mellitus Combined with Cerebral Stroke. J Clin Lab Anal. 2017; 31:e22033.

25. Singh KK, Chaubey G, Chen JY, Suravajhala P. Decoding SARS-CoV-2 Hijacking of Host Mitochondria in Pathogenesis of COVID-19. Am J Physiol-Cell Physiol. 2020. https://doi.org/10.1152/ajpcell.00224.2020 PMID: 32510973