COVID-19: Impact on linguistic and genetic isolates of India

Prajjval Pratap Singh, Prashanth Suravajhala, Chandana Basu Mallick, Rakesh Tamang, Ashutosh Kumar Rai, Pratheusa Machha, Royana Singh, Abhishek Pathak, Vijay Nath Mishra, Pankaj Shrivastava, Keshav K. Singh, Kumarasamy Thangaraj, and Gyaneshwer Chaubey

The rapid expansion of coronavirus SARS-CoV-2 has impacted various ethnic groups all over the world. The burden of infectious diseases including COVID-19 are generally reported to be higher for the Indigenous people. The historical knowledge have also suggested that the indigenous populations suffer more than the general populations in the pandemic. Recently, it has been reported that the indigenous groups of Brazil have been massively affected by COVID-19. Series of studies have shown that many of the indigenous communities reached at the verge of extinction due to this pandemic. Importantly, South Asia also has several indigenous and smaller communities, that are living in isolation. Till date, despite the two consecutive waves in India, there is no report on the impact of COVID-19 for indigenous tribes. Since smaller populations experiencing drift may have greater risk of such pandemic, we have analysed Runs of Homozygosity (ROH) among South Asian populations and identified several populations with longer homozygous segments. The longer runs of homozygosity at certain genomic regions may increases the susceptibility for COVID-19. Thus, we suggest extreme careful management of this pandemic among isolated populations of South Asia.

INTRODUCTION

It has been more than 18 months since the first case of COVID-19 was reported in India. Till now, India has been hit by two major waves with several hundred thousand death toll [1]. The devastating second wave was mainly driven by the alpha and delta variants [2, 3]. Researches on the delta variant have shown that it is more than twice as contagious as the Wuhan strain [4, 5]. Before the second wave in India, the third sero-survey conducted during months of December-January, has reported only 21.4% seropositivity [6]. Nevertheless, by the time the second wave arrived, a large number of seropositive people had exhausted the antibodies [7]. Perhaps, this gave an open field to alpha and delta virus variants to sweep. Otherwise, both of these variants were reported in India by the end of 2020, but their ruthless form was seen from April 2021 onwards [8]. This indicates that the waning antibody was the key driving force behind the second wave [7, 9], while the alpha and delta virus variants catalysed the intensity [3, 5, 8].

In the fourth serosurvey conducted by ICMR (Indian Council of Medical Research) during June-July 2021, the seropositivity have been found among 68% of the people [10], which is more than three times larger to third serosurvey [6]. The presence of antibodies in large number of people reflect the predominance of the second wave. The seropositivity in such a large population also suggests that a third major nationwide outbreak is unlikely in the recent past. The COVID-19 cases has been reported each and every regions of India, however, it is not known that how it has impacted the isolated and smaller populations [11].

With the global range expansion of coronavirus SARS-CoV-2, it is a matter of concern to protect vulnerable tribal populations from contagion. Various reports from Brazil have suggested that many of the indigenous communities were hard hit by the coronavirus SARS-CoV-2 [12–16]. India is a country of diverse endogamous tribal populations, speaking various languages [17]. Altogether, tribal populations make 8% of the total Indian census with some of major tribes e.g., Gond, Kol and Bhil populations, who are millions in number [18]. Yet, many of the South Asian tribal populations have experienced severe bottlenecks and are less than a thousand in numbers [19]. There has not been any study so far, on the impact of COVID-19 among these isolated and smaller populations.

In a broader demographic perspective, South Asia is a diverse place with hundreds of ethnolinguistic groups [20]. This is due to long term isolation, genetic drift and endogamy which collectively created unique genetic profile of South Asians [21]. Generally, high level of genetic diversity for a population implies high heterozygosity [22, 23]. This high level of genetic diversity is beneficial to populations for several reasons. In this context, when
the genes of individuals in a population vary greatly, it facilitates the populations for better fitness, including survival against infectious diseases [24, 25]. Thus in case of pandemics, the greater diversity in a population reduces the risk of extinction. The COVID-19 low case fatality rate among South Asians was likely due to multiple factors including genetics [26–28] and prior exposure to various pathogens [29, 30]. However, it is important to note that the East Asian-specific signal of positive selection against coronavirus has not been observed among South Asian populations [31].

We inherit every single copy of chromosomes from each of our parents. Our genome contains several homozygous segments or haplotypes where we receive identical or different copies from each of our parents. In consanguineous marriages, the chances of receiving identical copies are high [32]. These identical copies are also known as Runs of Homozygosity (ROH) [33, 34]. The genetic drift for a smaller population tends to increase the ROH. Studying ROH is important for understanding underlying levels of genetic variation [35]. ROH has been used extensively to study population structure, demographic history and genetic structure of complex diseases [33, 34]. It has been shown that the populations with longer ROH are enriched for deleterious variations [35–38]. Though, most of the South Asian populations carry a high level of genetic diversity, few genetic and linguistic isolates as well as historically migrated populations may have low effective population sizes (Ne) and experienced bottleneck and drift in the past. Hence, longer ROH carrying populations may have greater risk to ongoing pandemic. Here, we have studied the runs of Homozygosity (ROH) among South Asian populations and found out that many of the smaller and isolated populations have high number of long ROH segments.

MATERIALS AND METHODS
We have used publicly available datasets on Indian populations to estimate the Runs of Homozygosity (ROH) [39, 40]. PLINK 1.9 [41], was used for data management. The ROH for each of the populations was calculated using PLINK 1.9 [41]. We have used ‘-homog’ function to perform the analysis. For the calculations, we have used 1000 kb windows size with a minimum of 100 SNPs per window allowing one heterozygous and five missing calls per window. The designated window sequentially scans each and every individual and estimate for proportion in a homozygous window for every SNP.

RESULTS AND DISCUSSION
Among South Asian ethnic groups, majority of the populations have smaller and fewer number of ROH segments, whereas few of isolated as well as historically migrated populations carried ROH of longer and larger in size (Fig. 1). Historically migrated populations such as Parsis and Jews have their unique demographic history with smaller numbers of past effective population sizes (Ne) and follow strict endogamy [40, 42, 43]. These groups have migrated to South Asia in the last two millennia with limited founders. The molecular data further revealed that there was a sex-biased admixture with the local females followed by a high level of endogamy [40, 42, 43]. With their low level of heterozygosity these historically migrated populations may have a higher risk of COVID-19.

Among the studied groups, Andaman Islanders have the highest number as well as longest ROH segments (Fig. 1). Great Andamanese (census 43), Onge (census 100), Jarawa (census 375) and Sentinels (census 39) are the aboriginal tribal populations of these islands. Genetic studies on them [44–47] have suggested their deep rooted ancestry sharing with the South Asian, East Asian, Southeast Asian and Papuan populations [44–47]. It has been shown that much of the East and Southeast Asian populations are derived from the admixture of Andaman and Tiyuan [48] related ancestries [49]. Andamanese Islanders live in protected areas and the general public is not allowed to interact with them. However, seeing some of the past experiences [50] and number of cases at the Island among the general population, they are at greater risk, mainly from illegal intruders and health workers.

Studies have identified ACE2 as a host receptor for the SARS-CoV-2 [51, 52]. It has been shown that a polymorphism rs2285666 (G > A) of ACE2 gene of X chromosome may increase the expression level upto 50% [53–55]. This polymorphism was also widespread in South Asia and the haplotype associated with this SNP was shared with the East Eurasian populations [28]. Another SNP rs10490770 (T > C) at chromosome 3, introgressed from Neanderthal was also found to be associated with the disease severity mainly among European populations [56]. We have examined both of the SNPs with the Indian statewise infection and case fatality rates, and found a significant association of rs2285666 (but not for rs10490770) [57]. Interestingly, SNP rs2285666 (A) showed a clinal distribution with East and West Eurasia, whereas SNP 10490770 (C) had a frequent distribution primarily in the South Asia [57].

Looking at the clinal distribution of the SNP rs2285666, one may also argue its arrival to South Asia from East and Southeast Asia via geneflow [58, 59], and an isolation by distance (IBD) model for its present distribution. We agree that the spatial distribution of this SNP is significantly associated with the East/Southeast Asian-specific ancestry (R2 = 0.76; p = 9.44 × 10−6). Nevertheless, in comparison with the limited language associated (Austronesian and Tibeto-Burman), spatial distribution of East/Southeast Asian-specific ancestries [59–61], this SNP is much more frequent and widespread, well beyond the linguistic boundaries in South Asia [27, 57]. Moreover, a recent study on the hospital samples have reported a twofold increase for infection risk as well as threefold more chance of mortality with the risk allele rs2285666 (G) polymorphism [62].

In order to understand the susceptibility of isolated Andaman Islanders, we have estimated frequency of these SNPs among Jarawa and Onge populations. Notably, despite their closer genetic affinity with the ancestral East/Southeast Asian populations [45, 49], they have high frequency of risk allele (C) of rs10490770 (Jarawa 0.26 and Onge 0.29). Such high frequency of Neanderthal-specific allele adds an interesting aspect keeping in mind the 25KYA (Kilo Years Ago) split time with the South Asian populations [63]. For the ACE2 risk polymorphism rs2285666 (G), the Jarawa and Onge showed frequency of 0.58 and 0.35, respectively. If we compare Tibeto-Burman or Austrasian
isolated populations, e.g., language isolates—Nihali [64], genetic heterozygosity. 

Among the populations with the larger segments, it is pertinent that the population carrying segments larger than 150 Mb. In most of the isolated populations (isolated populations) with longer ROH may have higher susceptibility to SARS-CoV-2. Apart from these known isolated populations, we have also found out several Dravidian speaking groups harbouring high homozygosity (Fig 1). Interestingly, these Dravidian speakers with larger size and numbers of homozygous segments are from both tribal as well as caste populations. Among the populations other than Dravidian, carrying homozygous segments of more than 150 Mb, only a single group, each of Himalayan region (Changpa) and Austrasiatic (Kissan) are present among studied populations. Thus, majority of larger segments were present among Dravidian speakers. Interestingly, in the analysed dataset we did not find any Indo-European speaking population carrying segments larger than 150 Mb. In most of the populations with the larger segments, it is pertinent that the smaller population size and high level of inbreeding have reduced the heterozygosity.

In addition, with the studied populations, there are several isolated populations, e.g., language isolates—Nihali [64], genetic isolates—Abujhmaria [65], and many more who have shown the high ROH. Although, these populations are not well connected with the mainstream populations, however, there are high probabilities for them to contract with this virus seeing its nature of infectivity and range expansion. Furthermore, keeping in view of SARS-CoV-2 medical procedures, and lack of viable healthcare modern facilities, therefore, we suggest a high priority protection and utmost care for these isolated groups, so that we should not suffer to lose some of the living treasures of modern human evolution.

DATA AVAILABILITY

All datasets generated for this study are included in the article.

REFERENCES

1. Coronavirus in India: Latest Map and Case Count [Internet], (cited 2020 May 13).

2. Gupta N, Kaur H, Yadav PD, Mukhopadhayay L, Sahay RR, Kumar A, et al. Clinical Characterization and Genomic Analysis of Samples from COVID-19 Breakthrough Infections during the Second Wave among the Various States of India. Viruses. 2021;13:1782.

3. Ranjan R, Sharma A, Verma MK Characterization of the Second Wave of COVID-19 in India. medRxiv. 2021;https://doi.org/10.1101/2021.04.17.21255665.

4. Liu C, Ginn HM, Dejnirattisai W, Supasa P, Wang B, Tuekprakhon A, et al. Reduced neutralization of SARS-CoV-2 B. 1.617 by vaccine and convalescent serum. Cell. 2021;184:4220–36.

5. Roy B, Dhillon J, Habib N, Pugazhandhi B. Global variants of COVID-19: Current understanding. J Biomed Sci. 2021;8:8–11.

6. Murhekar MV, Bhatnagar T, Selvaraju S, Saravananukumar V, Thangaraj JWW, Shah N, et al. SARS-CoV-2 antibody seroprevalence in India, August–September, 2020: findings from the second nationwide household serosurvey. Lancet Glob Health. 2021;9:e257–66.

7. Singh PP, Chaubey G RE: Why is there a second wave in India? 2021; Science e-letter.

8. Kupferschmidt K, Wadman M. Delta variant triggers new phase in the pandemic. Science. 2021;372:1375–76.

9. Naushin S, Sardana V, Ujjainiya R, Bhatheja N, Kutum R, Bhaskar AK, et al. Insights from a Pan India Sero-Epidemiological survey (Phenome-India Cohort) for SARS-CoV2. Elife 2021:10:e66537.

10. Koshi J. Coronavirus | Kerala has the lowest seroprevalence among 21 States, M. P. has the highest: ICMR study. The Hindu [Internet]. 2021 Jul 28 (cited 2021 Aug 6); Available from: https://www.thehindu.com/news/national/coronavirus-kerala-has-the-lowest-seroprevalence-among-21-states-mp-has-the-highest-icmr-study/article35594182.ece

11. Power T, Wilson D, Best O, Brockie T, Bearskin LB, Millender E, et al. COVID-19 and indigenous peoples: an imperative for action. J Clin Nurs. 2020; https://doi.org/10.1111/jocn.15320.

12. Amigo I. Indigenous communities in Brazil fear pandemic’s impact. Science. 2020;368:362.

13. Charlier P, Varison L. Is COVID-19 being used as a weapon against Indigenous Peoples in Brazil? Lancet. 2020;396:1069–70.

14. Ferrante L, Feamside PM. Protect Indigenous peoples from COVID-19. Science. 2020;368:251–251.

15. Palamini CVC, Ortega MM, Marson FAL. COVID-19 in the Indigenous Population of Brazil. J Racial Ethn Health Disparities. 2020;7:1053–8.

16. Polidoro M, de Assis Mendonça F, Meneghel SN, Alves-Brito A, Gonzàlvex M, Bairro F, et al. Territories under siege: risks of the decimation of indigenous and Quilombolas peoples in the context of COVID-19 in South Brazil. J Racial Ethn Health Disparities. 2021;8:1119–1129.

17. Kivisild T, Rootsi S, Metspalu M, Parkk J, Kaldma K, et al. The genetics of language and farming spread in India. In: Bellwood P, Renfrew C, editors. The evolution of human society and language. London: Routledge; 2018. p. 37–66.

18. Census of India Website: Ofﬁce of the Registrar General & Census Commissioner, India [Internet]. (cited 2021 Jan 5). Available from: https://censusindia.gov.in/2011-common/censusdata2011.html.

19. Cooper Z. Archaeology and History: Early settlements in the Andaman Islands. New Delhi and Oxford: Oxford University press; 2002.

20. 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.

21. Chaubey G, Metspalu M, Kivisild T, Villems R. Peopling of South Asia: investigating the caste-tribe continuum in India. BioEssays. 2007;29:91–100.

22. Reed DH, Frankham R. Correlation between ﬁtness and genetic diversity. Conserv Biol. 2003;17:230–7.

23. Tishkoff SA, Verrelli BC. Patterns of human genetic diversity: implications for human evolutionary history and disease. Annu Rev Genomics Hum Genet. 2003;4:293–340.

24. Lyons EJ, Frodsham AJ, Zhang L, Hill AV, Amos W. Consanguinity and susceptibility to infectious diseases in humans. Biol Lett. 2009;5:574–6.

25. Cooke GS, Hill AV. Genetics of susceptibility to human infectious disease. Nat Rev Genet. 2001;2:967–77.

26. Chaubey G, Coronavirus (SARS-CoV-2) and Mortality Rate in India: The Winning Edge, Front Public Health. 2020:8:397.

27. Srivastava A, Bandopadhyay A, Das D, Pandey RK, Singh V, Khanam N, et al. Genetic association of ACE2 rs2285666 polymorphism with COVID-19 spatial distribution in India. Front Genet. 2020;11:1163.

28. Srivastava A, Pandey RK, Singh PP, Kumar P, Rasalanka AA, Tamang R, et al. Most frequent South Asian haplotypes of ACE2 share identity by descent with East Eurasian populations. PLoS ONE. 2020;15(Sep):e0238255.

29. Mukherjee S, Sarkar-Roy N, Wagener DK, Majumder PP. Signatures of natural selection with the low seroprevalence among 21 states—mp has the highest: ICMR study. The Hindu [Internet]. 2021 Jul 28 (cited 2021 Aug 6); Available from: https://www.thehindu.com/news/national/coronavirus-kerala-has-the-lowest-seroprevalence-among-21-states-mp-has-the-highest-icmr-study/article35594182.ece

30. Soulimy Y, Lauterberg ME, Tobler R, Huber CD, Johar AS, Moradi SV, et al. The genetic basis for the low seroprevalence among 21 states—mp has the highest: ICMR study. The Hindu [Internet]. 2021 Jul 28 (cited 2021 Aug 6); Available from: https://www.thehindu.com/news/national/coronavirus-kerala-has-the-lowest-seroprevalence-among-21-states-mp-has-the-highest-icmr-study/article35594182.ece

31. Soulimy Y, Lauterberg ME, Tobler R, Huber CD, Johar AS, Moradi SV, et al. The genetic basis for the low seroprevalence among 21 states—mp has the highest: ICMR study. The Hindu [Internet]. 2021 Jul 28 (cited 2021 Aug 6); Available from: https://www.thehindu.com/news/national/coronavirus-kerala-has-the-lowest-seroprevalence-among-21-states-mp-has-the-highest-icmr-study/article35594182.ece

32. Hamamy H, Antonarakis SE, Cavalli-Sforza LL, Temtamy S, Romeo G, Ten Kate LP, et al. The genetics of language and farming spread in India. In: Bellwood P, Renfrew C, editors. The evolution of human society and language. London: Routledge; 2018. p. 37–66.

33. Srivastava A, Bandopadhyay A, Das D, Pandey RK, Singh V, Khanam N, et al. Genetic association of ACE2 rs2285666 polymorphism with COVID-19 spatial distribution in India. Front Genet. 2020;11:1163.
The major genetic risk factor for severe COVID-19 does not show any association among South Asian populations. Sci Rep. 2021;11:1

The promise of discovering population-specific disease-associated genes in South Asia. Nat Genet. 2017;49:1403.

Pathak AK, Srivastava A, Singh PP, Das D, Bandopadhayay A, Singh P, et al. Histric migration to South Asia in the last two millennia: A case of Jewish and Parsi populations. J Biosci. 2019;44(Jul):72.

Chang CC, Chow CC, Tellier LC, Vattikuti S, Purcell SM, Lee JJ Second-generation PLINK: rising to the challenge of larger and richer data sets. GigaScience. 2015;4:13742-015-0047-8.

Chaubey G, Endicott P. The Andaman Islanders in a regional genetic context: reexamining the evidence for an early peopling of the archipelago from South Asia. Hum Biol. 2013;85:153–72.

Aghahadian F, Yunus Y, Naidu R, Jinam T, Manica A, Hoh BP, et al. Unraveling the genetic history of negritos and indigenous populations of southeastern Asia. Genome Biol. 2015;17:1206–15.

Basu A, Sarkar-Roy N, Majumder PM. Genomic reconstruction of the history of the Jewish populations of India. Sci Rep. 2016;6:19166.

Thangaraj K, Gupta NJ, Pavani K, Reddy AG, Subramainan S, Rani DS, et al. Y chromosome deletions in azoospermic men in India. J Androl. 2003;24:588–97.

Chaubey G, Endicott P. The Andaman Islanders in a regional genetic context: reexamining the evidence for an early peopling of the archipelago from South Asia. Hum Biol. 2013;85:153–72.

Aghahadian F, Yunus Y, Naidu R, Jinam T, Manica A, Hoh BP, et al. Unraveling the genetic history of negritos and indigenous populations of southeastern Asia. Genome Biol. 2015;17:1206–15.

Basu A, Sarkar-Roy N, Majumder PM. Genomic reconstruction of the history of the Jewish populations of India. Sci Rep. 2016;6:19166.

Thangaraj K, Gupta NJ, Pavani K, Reddy AG, Subramainan S, Rani DS, et al. Y chromosome deletions in azoospermic men in India. J Androl. 2003;24:588–97.