

## BRIEF COMMUNICATION OPEN



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

Prajval Pratap Singh<sup>1</sup>, Prashanth Suravajhala<sup>2,3</sup>, Chandana Basu Mallick<sup>4</sup>, Rakesh Tamang<sup>5</sup>, Ashutosh Kumar Rai<sup>6</sup>, Pratheusa Machha<sup>7,8</sup>, Royana Singh<sup>9</sup>, Abhishek Pathak<sup>10</sup>, Vijay Nath Mishra<sup>10</sup>, Pankaj Shrivastava<sup>11</sup>, Keshav K. Singh<sup>12</sup>, Kumarasamy Thangaraj<sup>7,13</sup> and Gyaneshwer Chaubey<sup>1</sup>

© The Author(s), under exclusive licence to Springer Nature Limited 2021

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.

*Genes & Immunity* (2022) 23:47–50; <https://doi.org/10.1038/s41435-021-00150-8>

## 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 tribals 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

<sup>1</sup>Cytogenetics Laboratory, Department of Zoology, Banaras Hindu University, Varanasi 221005, India. <sup>2</sup>Department of Biotechnology and Bioinformatics, Birla Institute of Scientific Research Statue Circle, Jaipur, Rajasthan, India. <sup>3</sup>Amrita School of Biotechnology, Amrita University Kerala India, Vallikavu 690525, India. <sup>4</sup>Centre for Genetic Disorders, Institute of Science, Banaras Hindu University, Varanasi 221005, India. <sup>5</sup>Department of Zoology, University of Calcutta, Kolkata 700019, India. <sup>6</sup>Department of Biochemistry, College of Medicine, Imam Abdulrahman Bin Faisal University, Dammam, Saudi Arabia. <sup>7</sup>CSIR-Centre for Cellular and Molecular Biology, Hyderabad 500007, India. <sup>8</sup>Academy of Scientific and Innovative Research, (AcSIR), Ghaziabad 201002, India. <sup>9</sup>Department of Anatomy, Institute of Medical Sciences, Banaras Hindu University, Varanasi 221005, India. <sup>10</sup>Department of Neurology, Institute of Medical Sciences, Banaras Hindu University, Varanasi 221005, India. <sup>11</sup>Department of Home (Police), DNA Fingerprinting Unit, State Forensic Science Laboratory, Government of MP, Sagar, India. <sup>12</sup>Department of Genetics, School of Medicine, University of Alabama at Birmingham, Kaul Genetics Building, Birmingham, AL, USA. <sup>13</sup>Centre for DNA Fingerprinting and Diagnostics (CDFD), Hyderabad 500039, India. ✉email: thangs@ccmb.res.in; gyaneshwer.chaubey@bhu.ac.in

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 ( $N_e$ ) 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.

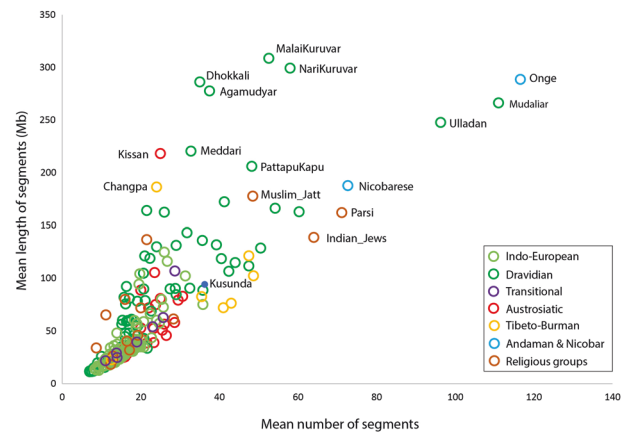
## MATERIALS AND METHODS

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 ‘-homozyg’ 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 ( $N_e$ ) 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 (the genetic study of Sentinels have not been done yet), 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 Tianyuan [48] related



**Fig. 1 Two dimensional plot of mean number (X-axis) vs mean length (Mb) (Y-axis) of homozygous chromosomal segments among various South Asian populations.** Populations with higher than 150 Mb chromosomal segments as well as linguistic isolates have been marked.

ancestries [49]. Andaman 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 gene flow [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 ( $R^2 = 0.76$ ;  $p = 9.44 \times 10^{-6}$ ). Nevertheless, in comparison with the limited language associated (Austroasiatic 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 Austroasiatic

**Table 1.** The frequency distribution (95%CI) among Austroasiatic, Tibeto-Burman and Andamanese groups for the risk alleles observed in the present study.

Population group	N	rs10490770(C)	rs2285666(G)
Austroasiatic	106	0.302 (0.223–0.395)	0.274 (0.198–0.366)
Tibeto-Burman	96	0.063 (0.030–0.130)	0.135 (0.081–0.218)
Jarawa	19	0.263 (0.119–0.491)	0.579 (0.361–0.769)
Onge	17	0.294 (0.133–0.535)	0.353 (0.173–0.590)

populations (with relatively smaller ROH segments) analysed in the present study, they always tend to show significantly (two tailed  $p < 0.001$ ), lower frequency of the risk alleles for SNP rs2285666 (G) (Table 1). Thus, here in case of Andaman Islands (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 large 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 Austroasiatic (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]. Available from: <https://www.covid19india.org>
2. Gupta N, Kaur H, Yadav PD, Mukhopadhyay 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, Saravanakumar V, Thangaraj JWV, 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 there is 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. Koshy 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;368:352.
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, Fearnside PM. Protect Indigenous peoples from COVID-19. *Science*. 2020;368:251–251.
15. Palamim 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, Gonçalves M, Bairros 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, Metspalu E, Parik J, Kaldma K, et al. The genetics of language and farming spread in India. In: Bellwood P, Renfrew C, editors. *Examining the farming/language dispersal hypothesis*. Cambridge: The McDonald Institute for Archaeological Research; 2003. p. 215–22.
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 N. Rev Mol Cell Dev Biol*. 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, Rasalkar 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 are not uniform across genes of innate immune system, but purifying selection is the dominant signature. *Proc Natl Acad Sci*. 2009;106:7073–8.
30. Bairagya BB, Bhattacharya P, Bhattacharya SK, Dey B, Dey U, Ghosh T, et al. Genetic variation and haplotype structures of innate immunity genes in eastern India. *Infect Genet Evol*. 2008;8:360–6.
31. Soulimi Y, Lauterbur ME, Tobler R, Huber CD, Johar AS, Moradi SV, et al. An ancient viral epidemic involving host coronavirus interacting genes more than 20,000 years ago in East Asia. *Curr Biol*. 2021;31:3504–3514.
32. Hamamy H, Antonarakis SE, Cavalli-Sforza LL, Teamtamy S, Romeo G, Ten Kate LP, et al. Consanguineous marriages, pearls and perils: Geneva international consanguinity workshop report. *Genet Med*. 2011;13:841–7.
33. McQuillan R, Leutenegger A-L, Abdel-Rahman R, Franklin CS, Pericic M, Barac-Lauc L, et al. Runs of homozygosity in European populations. *Am J Hum Genet*. 2008;83:359–72.
34. Kirin M, McQuillan R, Franklin CS, Campbell H, McKeigue PM, Wilson JF. Genomic runs of homozygosity record population history and consanguinity. *PloS One*. 2010;5:e13996.

35. Ceballos FC, Joshi PK, Clark DW, Ramsay M, Wilson JF. Runs of homozygosity: windows into population history and trait architecture. *Nat Rev Genet.* 2018;19:220–34.
36. Christofidou P, Nelson CP, Nikpay M, Qu L, Li M, Loley C, et al. Runs of homozygosity: association with coronary artery disease and gene expression in monocytes and macrophages. *Am J Hum Genet.* 2015;97:228–37.
37. Ghani M, Reitz C, Cheng R, Vardarajan BN, Jun G, Sato C, et al. Association of long runs of homozygosity with Alzheimer disease among African American individuals. *JAMA Neurol.* 2015;72:1313–23.
38. Szpiech ZA, Xu J, Pemberton TJ, Peng W, Zöllner S, Rosenberg NA, et al. Long runs of homozygosity are enriched for deleterious variation. *Am J Hum Genet.* 2013;93:90–102.
39. 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.
40. Pathak AK, Srivastava A, Singh PP, Das D, Bandopadhyay A, Singh P, et al. Historic migration to South Asia in the last two millennia: A case of Jewish and Parsi populations. *J Biosci.* 2019;44(Jul):72.
41. 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: s13742-015-0047-8.
42. Chaubey G, Ayub Q, Rai N, Prakash S, Mushrif-Tripathy V, Mezzavilla M, et al. "Like sugar in milk": reconstructing the genetic history of the Parsi population. *Genome Biol.* 2017;18:110.
43. Chaubey G, Singh M, Rai N, Kariappa M, Singh K, Singh A, et al. Genetic affinities of the Jewish populations of India. *Sci Rep.* 2016;6:19166.
44. Thangaraj K, Gupta NJ, Pavan K, Reddy AG, Subramanian S, Rani DS, et al. Y chromosome deletions in azoospermic men in India. *J Androl.* 2003;24:588–97.
45. 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.
46. Aghakhanian F, Yunus Y, Naidu R, Jinam T, Manica A, Hoh BP, et al. Unravelling the genetic history of negritos and indigenous populations of southeast Asia. *Genome Biol Evol.* 2015;7:1206–15.
47. Basu A, Sarkar-Roy N, Majumder PP. Genomic reconstruction of the history of extant populations of India reveals five distinct ancestral components and a complex structure. *Proc Natl Acad Sci USA.* 2016;113:1594–9.
48. Fu Q, Meyer M, Gao X, Stenzel U, Burbano HA, Kelso J, et al. DNA analysis of an early modern human from Tianyuan Cave, China. *Proc Natl Acad Sci.* 2013;110:2223–7.
49. Wang C-C, Yeh H-Y, Popov AN, Zhang H-Q, Matsumura H, Sirak K, et al. Genomic insights into the formation of human populations in East Asia. *Nature.* 2021;591:413–9.
50. Ghosh, S. Why the Andaman tribes need isolation [Internet]. *Nature India.* [cited 2020 Jul 2]. Available from: <https://www.natureasia.com/en/nindia/article/10.1038/nindia.2019.39>
51. 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. *Lancet.* 2020;395:565–74.
52. 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.
53. 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.
54. 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. *Aging.* 2020;12:10087.
55. Chen Y, Zhang P, Zhou X, Liu D, Zhong J, Zhang C, et al. Relationship between genetic variants of ACE 2 gene and circulating levels of ACE 2 and its metabolites. *J Clin Pharm Ther.* 2018;43:189–95.
56. Zeberg H, Pääbo S. The major genetic risk factor for severe COVID-19 is inherited from Neanderthals. *Nature.* 2020;587:610–2.
57. Singh PP, Srivastava A, Sultana GNN, Khanam N, Pathak A, Suravajhala P, et al. The major genetic risk factor for severe COVID-19 does not show any association among South Asian populations. *Sci Rep.* 2021;11:1–4.
58. 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.
59. Tamang R, Chaubey G, Nandan A, Govindaraj P, Singh VK, Rai N, et al. Reconstructing the demographic history of the Himalayan and adjoining populations. *Hum Genet.* 2018;137:129–39.
60. Thangaraj K, Chaubey G, Kivisild T, Selvi Rani D, Singh VK, Ismail T, et al. Maternal footprints of Southeast Asians in North India. *Hum Hered.* 2008;66:1–9.
61. Chaubey G. East Asian Ancestry in India. *Ind J Phys Anthr Hum Genet.* 2015;34:193–9.
62. Möhlendick B, Schönfelder K, Breuckmann K, Elsner C, Babel N, Balfanz P, et al. ACE2 polymorphism and susceptibility for SARS-CoV-2 infection and severity of COVID-19. *Pharmacogenet Genomics.* 2021; <https://doi.org/10.1097/FPC.0000000000000436>
63. Barik SS, Sahani R, Prasad BV, Endicott P, Metspalu M, Sarkar BN, et al. Detailed mtDNA genotypes permit a reassessment of the settlement and population structure of the Andaman Islands. *Am J Phys Anthr.* 2008;136:19–27.
64. Nagaraja, KS. *The Nihali Language (Grammar, Texts and Vocabulary).* Mysore: Central Institute of Indian Languages; 2014.
65. GenomeAsia100K Consortium. The GenomeAsia 100K Project enables genetic discoveries across Asia. *Nature.* 2019;576:106.

## ACKNOWLEDGEMENTS

We thank Prof. George van Driem for his constructive comments. This work is supported by the Faculty IOE grant BHU (6031). RT and GC are supported by SERB India (CRG/2018/001727), PPS is supported by CSIR fellowship, CBM is supported by Wellcome Trust/DBT India Alliance Early Career Fellowship (IA/E/18/1/504338) and KT is supported by Council of Scientific and Industrial Research (CSIR) and J C Bose Fellowship (JCB/2019/000027) from Science and Engineering Research Board (SERB), Department of Science and Technology, Government of India.

## AUTHOR CONTRIBUTIONS

GC and KT conceived and designed this study. PPS, PrS, CBM, RT, AKR, PM, RS, AP, VNM, PS and KKS collected the data for studied alleles and COVID-19. PPS, PrS, and GC analyzed the runs of homozygosity data. PPS, PrS, KT, and GC wrote the manuscript from the inputs of other co-authors. All authors contributed to the article and approved the submitted version.

## COMPETING INTERESTS

The authors declare no competing interests.

## ADDITIONAL INFORMATION

**Correspondence** and requests for materials should be addressed to Kumarasamy Thangaraj or Gyaneshwer Chaubey.

**Reprints and permission information** is available at <http://www.nature.com/reprints>

**Publisher's note** Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



**Open Access** This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.

© The Author(s), under exclusive licence to Springer Nature Limited 2021