

1 **Investigating the likely association between genetic ancestry and COVID-19**

2 **manifestations**

3 Ranajit Das* and Sudeep D. Ghate

4 Yenepoya Research Centre, Yenepoya (Deemed to be University), Mangalore, Karnataka,

5 India.

6

7

8 * All correspondence to

9 Dr. Ranajit Das

10 Yenepoya Research Centre

11 Yenepoya (Deemed to be University)

12 University Road, Deralakatte

13 Mangalore 575018

14 Karnataka, India

15

16

17

18

19

20

21

22

23

24

25

26

27

28

29

30

31

32

33

34

35 **Abstract**

36

37 **Background**

38 The novel coronavirus: severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has
39 spread rapidly throughout the world leading to catastrophic consequences. However, SARS-
40 CoV-2 infection has shown discernible variability across the globe. While in some countries
41 people are recovering relatively quickly, in others, recovery times have been comparatively
42 longer and number of individuals succumbing to it are high. This variability in coronavirus
43 disease 2019 (COVID-19) susceptibility is suggestive of a likely association between the
44 genetic-make up of affected individuals modulated by their ancestry and the severity of
45 COVID-19 manifestations.

46 **Objective**

47 In this study, we aimed to evaluate the potential association between an individual's genetic
48 ancestry and the extent of COVID-19 disease presentation employing Europeans as the case
49 study. In addition, using a genome wide association (GWAS) approach we sought to discern
50 the putative single nucleotide polymorphism (SNP) markers and genes that may be likely
51 associated with differential COVID-19 manifestations by comparative analyses of the
52 European and East Asian genomes.

53 **Method**

54 To this end, we employed 10,215 ancient and modern genomes across the globe assessing
55 597,573 SNPs obtained from the databank of Dr. David Reich, Harvard Medical School,
56 USA to evaluate the likely correlation between European ancestry and COVID-19
57 manifestations. Ancestry proportions were determined using qpAdm program implemented in
58 AdmixTools v5.1. Pearson's correlation coefficient (r) between various ancestry proportions
59 of European genomes and COVID-19 death/recovery ratio was calculated and its significance
60 was statistically evaluated. Genome wide association study (GWAS) was performed in
61 PLINK v1.9 to investigate SNPs with significant allele frequency variations among European
62 and East Asian genomes that likely correlated with differential COVID-19 infectivity.

63 **Results**

64 We found significant positive correlation ($r=0.58$, $P=0.03$) between West European hunter
65 gatherers (WHG) ancestral fractions and COVID-19 death/recovery ratio for data as of 5th
66 April 2020. This association discernibly amplified ($r=0.77$, $P=0.009$) upon reanalyses based
67 on data as of 30th June 2020, removing countries with small sample sizes and adding those
68 that are a bridge between Europe and Asia. Using GWAS we further identified 404 immune

69 response related SNPs by comparing publicly available 753 genomes from various European
70 countries against 838 genomes from various Eastern Asian countries. Prominently, we
71 identified that SNPs associated with immune-system related pathways such as interferon
72 stimulated antiviral response, adaptive and innate immune system and IL-6 dependent
73 immune responses show significant differences in allele frequencies [Chi square values
74 (≥ 1500 ; $P \approx 0$)] between Europeans and East Asians.

75 Conclusion

76 So far, to the best of our knowledge, this is the first study investigating the likely association
77 between host genetic ancestry and COVID-19 severity. These findings improve our overall
78 understanding of the putative genetic modifiers of COVID-19 clinical presentation. We note
79 that the development of effective therapeutics will benefit immensely from more detailed
80 analyses of individual genomic sequence data from COVID-19 patients of varied ancestries.

81

82 **Key words:** Genomic ancestry; COVID-19 host genetics; COVID-19 manifestations; novel
83 coronavirus; Genome Wide Association study (GWAS)

84

85 Introduction

86

87 The novel coronavirus, severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), was
88 officially identified on 7th January 2020 in Wuhan, China [1]. Since then, it has spread
89 rapidly throughout the world leading to devastating consequences. Notably, COVID-19,
90 caused by SARS-CoV-2 has shown significant worldwide variability, especially in terms of
91 the death/recovery ratio that pertains to a ratio of the number of deaths caused by it to
92 numbers of infected individuals who have recovered within the same time interval. While
93 affected individuals in some parts of the world have recovered relatively quickly with lower
94 morbidities, those from various other regions appear to remain sick for longer with slower
95 recovery times, more debilitating consequences and demonstrate relatively higher
96 death/recovery ratios [2]. For instance, Verity *et al.* (2020) estimated an overall fatality ratio
97 of 2.7% for COVID-19 cases reported outside of mainland China. For people with travel
98 history to mainland China this ratio was 1.1% and those without any travel history to China
99 the ratio was 3.6% [3]. This is suggestive of a likely population level genetic variation in
100 terms of susceptibility to SARS-CoV-2 and COVID-19 manifestations.

101 A case in point in this context appears to be *ACE2* that encodes for angiotensin I
102 converting enzyme 2 demonstrated to be the host receptor for the SARS-CoV-2 [4,5].
103 Recently, Cao *et al.* analysed 1700 variants in *ACE2* from The China Metabolic Analytics
104 Project (ChinaMAP) and 1000 Genomes Project databases [6]. Their study revealed
105 significant variation in allele frequencies among the populations assessed, notably among
106 people of East Asian and European ancestries. They speculated that the differences in allele
107 frequencies of *ACE2* coding variants, likely associated with its elevated expression may
108 influence *ACE2* function and putatively impact the susceptibility, symptoms, and outcome of
109 SARS-CoV-2 infection in evaluated populations. Additionally, a majority of the 15
110 expression quantitative trait loci (eQTL) variants they identified had discernibly higher allele
111 frequencies among East Asian populations compared to Europeans, which may suggest
112 differential susceptibility towards SARS-CoV-2 in these two populations under similar
113 conditions. Stawiski *et al.* who performed a comprehensive analysis on >290,000 samples
114 representing > 400 populations, also identified natural *ACE2* variants that can potentially
115 alter the virus-host interaction and thus alter host susceptibility towards COVID-19
116 manifestations [7]. While some of these variants can potentially increase susceptibility for
117 COVID-19, others were speculated to be protective and surmised to display decreased
118 binding to SARS-CoV-2 spike protein and thus decrease susceptibility to infection.

119 In addition to *ACE2*, COVID-19 manifestations may also be modulated by several
120 immune response modulating genes. Li *et al.* reviewed the interaction between coronaviruses
121 and the innate immune systems of the hosts [8]. It was speculated that coronavirus infection
122 can result in secretion of higher amounts of chemokines and cytokines such as interleukins
123 (IL-1, IL-6, IL-8, IL-21), Tumor necrosis factor (TNF)- β and monocyte chemoattractant
124 protein-1 (MCP-1) in infected cells. Further, the hosts' major antiviral machinery, largely
125 composed of interferons (IFNs) can also act to restrain SARS-CoV-2 infection. Overall it
126 may be speculated that variations in host genes involved in its antiviral response may be
127 responsible for differential clinical presentation of COVID-19. Recently, genetic variability
128 across major histocompatibility complex (MHC) class I genes [9] and apolipoprotein E
129 (*APoE*) [10] have also been associated with severity of COVID-19 manifestations.

130 Further recent studies have shown that variants associated with immune
131 responsiveness display significant population-specific signals of natural selection [11]. In
132 particular, the authors showed that admixture with Neanderthals may have differentially
133 shaped the immune systems in European populations, introducing novel regulatory variants,
134 which may affect preferential responses to viral infections. European genomes were further

135 modulated and distinguished owing to multiple waves of migration throughout their ancient
136 past. We note here that while Neolithic European populations predominantly descended from
137 Anatolian migrants, European ancestry shows a distinct West–East cline in indigenous West
138 European hunter gatherers (WHGs) [12]. At the same time Eastern Europe was in the
139 rudimentary stages of the formation of Bronze Age steppe ancestry, which later spread into
140 Central and Northern Europe through East-West expansion [13]. Further, high genetic
141 substructure between North-Western and South-Eastern hunter-gatherers has also been
142 documented recently [14].

143 Ancestry specific variation in immune responsiveness and its potential genetic and
144 evolutionary determinants are poorly understood especially in the context of SARS-CoV-2
145 infection and COVID-19 disease manifestations. In addition, despite its broad implications in
146 identifying novel variants putatively associated with variation in host immune responses
147 across different ancestries, only a handful of genome wide association study (GWAS) have
148 been performed pertaining to viral infections, especially in the context of SARS-CoV-2
149 infection [15].

150 Here we aimed to unravel the potential association between host genetic ancestry and
151 COVID-19 disease employing Europeans as the case study. In addition, using a GWAS
152 approach we sought to discern the putative single nucleotide polymorphism (SNP)s, genes
153 and pathways that may be potentially associated with differential COVID-19 infection by
154 comparative analyses of the European and East Asian genomes. To the best of our
155 knowledge, so far this is the first study to assess the relationship between host genetics and
156 SARS-CoV-2 severity comparing individuals from distinct ancestries.

157

158 **Method**

159

160 **Dataset**

161

162 Genome data for the current study was obtained from the personal database of Dr. David
163 Reich’s Lab Harvard Medical School, USA [16]. The final dataset comprised of 10,215
164 ancient and modern genomes across the globe assessing 597,573 single nucleotide
165 polymorphisms (SNPs). File conversions and manipulations were performed using
166 EIGENSTRAT (EIG) v7.2 [17] and PLINK v1.9 [18].

167 COVID-19 data was obtained from 2019 Novel Coronavirus COVID-19 Data
168 Repository by The Center for Systems Science and Engineering, Johns Hopkins (JHU-CSSE)

169 [2]. The data available in this web portal as of Sunday, 5th April 2020, 6 PM was used in the
170 primary analyses. To evaluate the robustness of our results, we repeated our analyses with the
171 data available in Johns Hopkins CSSE web portal as of 30th June 2020, 4:33 AM. Indian
172 state-wise data was obtained from John Hopkins CSSE portal and confirmed through
173 Government of India coronavirus portal as of 30th June 2020 [19].

174

175 Ancestry proportions in European genomes and COVID-19 manifestations

176

177 We employed whole genome data from ten European countries in the initial analyses (Table
178 1). During reanalysis, countries such as Norway, Czech Republic and Ukraine were removed
179 due to smaller sample size (N<15). Instead, Eastern European countries such as Romania,
180 Bulgaria, and those that bridge between Europe and Asia such as Georgia, Turkey and
181 Armenia were included. Further, UK that did not provide officially updated COVID-19
182 recovery information was excluded during reanalysis. Furthermore, a larger sample size was
183 employed for Italians (TSI) obtained from 1000 genomes project database (Table 2).

184 We used *qpAdm* [20] implemented in AdmixTools v5.1 [21,22] to estimate ancestry
185 proportions in the European genomes originating from a mixture of ‘reference’ populations
186 by utilizing shared genetic drift with a set of ‘outgroup’ populations. 14 ancient genomes
187 namely *Luxembourg_Loschbour.DG*, *Luxembourg_Loschbour_published.DG*,
188 *Luxembourg_Loschbour*, *Iberia_HG* (N=5), *Iberia_HG_lc*, *Iberia_HG_published*,
189 *LaBrana1_published.SG*, *Hungary_EN_HG_Koros* (N=2),
190 *Hungary_EN_HG_Koros_published.SG* were grouped together as West European Hunter-
191 Gatherers (WHGs); three ancient genomes namely *Russia_EHG*, *Russia_HG_Samara* and
192 *Russia_HG_Karelia* (N=2) were grouped together as Mesolithic hunter-gatherers of Eastern
193 Europe (EHGs); two ancient genomes: *KK1.SG*, *SATP.SG* were grouped as Caucasus hunter-
194 gatherers (CHGs) and *Sweden_HG_Motala* (N=8) were renamed as Scandinavian Hunter-
195 Gatherers (SHGs) in *qpAdm* analysis. We tried to model modern-day Europeans with
196 different combinations of Neolithic Near-East populations such as Neolithic Iranians
197 (*Iran_GanjDareh_N*), Neolithic Anatolians (*Anatolia_N*) and Natufians (*Israel_Natufian*)
198 alongside various combinations of ancient European Hunter Gatherer populations (WHG,
199 EHG, CHG and SHG) as mentioned in Lazaridis et al. (2016) [23]. We inferred that modern-
200 day Europeans could be best modelled as a combination of three source populations namely
201 EHG, WHGs and Neolithic Iranians (*Iran_GanjDareh_N*) as *Left* (EHG, WHG,
202 *Iran_GanjDareh_N*). We used a mixture of eight ancient and modern-day populations

203 comprising of *Ust Ishim, MAI, Kostenki14, Han, Papuan, Chukchi, Karitiana, Mbutiasour*
204 *'Right'* outgroup populations (O8).

205 Pearson's correlation coefficient (r) between various ancestry proportions of
206 European genomes and COVID-19 death/recovery ratio was calculated and its significance
207 was statistically evaluated using GraphPad Prism v8.4.0, GraphPad Software, San Diego,
208 California USA [24]. Additionally, we developed several linear regression models with
209 different combinations of WHG ancestry fraction (x_1) alongside climatic factors such as 15
210 days' mean temperatures (x_2) and 15 days' mean humidity (x_3) of the ten European countries
211 employed in this study to statistically evaluate their impact on COVID-19 death/recovery
212 ratio (y) using *glm* function implemented in R v3.6.3 (Table 1, Supplemental Table 1). One-
213 tailed tests were performed considering the alternative hypothesis of positive association of
214 x_1 , x_2 and x_3 with y ($H_A: \beta > 0$). The best model was determined by likelihood ratio test
215 (LRT), employing *lrtest* function, implemented in R package '*epicalc*'. Temperature and
216 humidity data were obtained from Time and Date.com [25], which garners meteorological
217 data from leading meteorological institutes such as The World Meteorological Organization
218 [26] and Meteorological Assimilation Data Ingest System (MADIS) [27]. Ancestry and
219 COVID-19 death/recovery ratio maps were mapped by country using QGIS version 3.12.3
220 [28].

221 Notably, the death/recovery ratio was used in this study as the test statistic as it can
222 potentially best reflect the extent to which individuals have succumbed versus those
223 recovering owing to COVID-19 infection. We believe that this ratio likely best reflects host
224 immune responsiveness against viral assault, especially since it does not take into account the
225 number of active cases and statistical biases arising due to population size, spread of the
226 disease and under-reporting/under-testing of COVID-19 cases.

227

228 Genome-wide association analyses (GWAS)

229

230 Genome wide association study (GWAS) was performed to identify novel genetic variants
231 that show significant variation between genomes of European and East Asian ancestries. In
232 GWAS, European genomes (N=753) with higher death/recovery ratio (Cases) were compared
233 against Eastern Asian genomes (N=838) with low death/recovery ratio (Controls). To this
234 end, standard case-control-based association analyses was performed in PLINK v1.9 using --
235 assoc command. A sorted list of association results will be generated using --adjust flag to the
236 --assoc command. Chi square test, implemented in PLINK --assoc command, was performed

237 separately for all 597,573 SNPs to statistically assess their significance and for stringency P
238 <0.01 was considered significant. A Manhattan plot was generated in Haploview [29] by
239 plotting Chi square values of all assessed SNPs to identify the SNPs that are likely associated
240 with COVID-19 manifestations. Highly significant SNPs were annotated using SNPnexus
241 web-based server [30].

242

243 **Results**

244

245 Ancestry proportions in European genomes and COVID-19 manifestations

246

247 We modelled all Europeans as a combination of three source populations namely EHGs,
248 WHGs and Neolithic Iranians (see Methods) in *qpAdm* analysis. Among 10 European
249 populations employed in this study, GBR (British in England and Scotland) genomes were
250 found to have the highest WHG ancestry proportions (26.1%) (Fig. 1a, Table 1). Notably,
251 among the European populations evaluated, COVID-19 death/recovery ratio till 5th April
252 2020, was the highest among the British people: death/recovery ratio=31.94 (Fig. 1b, Table
253 1). However, we omitted GBR genomes in our reanalysis due to absence of updated COVID-
254 19 recovery information from UK. In contrast, Russian and Finnish populations which could
255 not be modelled with WHG ancestry so far appear to have less detrimental COVID-19
256 manifestations with 0.02 and 0.05 death/recovery ratio, respectively (Fig. 1a and 1b).
257 Consistent with this, we found significant positive correlation (Pearson's Correlation; $r=0.58$;
258 $P=0.03$) between WHG ancestry fraction and COVID-19 death/recovery ratio. The
259 association discernibly increased ($r=0.77$, $P=0.009$) upon reanalysing the data, omitting
260 countries with small sample sizes and adding those that form a bridge between Europe and
261 Asia.

262 Congruent with correlation results, we found significant linear positive association
263 between WHG ancestry fraction and COVID-19 death/recovery ratio (Model 0) ($R^2=0.34$, P
264 $=0.039$) (Table 1). However, we did not find any association between 15 days' mean
265 temperature (Model 1) and humidity (Model 2) with COVID-19 death/recovery ratio (P
266 $=0.49$ and 0.29 respectively) (Table 1). Moreover, we found significant association between
267 the combination of WHG ancestry and temperature (Model 3), WHG ancestry and humidity
268 (Model 4), and WHG ancestry and a combination of both climatic factors (temperature and
269 humidity, Model 6) with COVID-19 death/recovery ratio ($R^2=0.33$, 0.39 and 0.4
270 respectively) (Table 1). However, in all cases the significance of the test was driven by the

271 WHG ancestry fraction alone ($P=0.04$, 0.04 and 0.05 for Model 3, Model 4 and Model 6
272 respectively) and not by the climatic factors ($P_{\text{temperature}}=0.54$, $P_{\text{humidity}}=0.23$, and
273 $P_{\text{temperature}}=0.61$ and $P_{\text{humidity}}=0.23$ for Model 3, Model 4 and Model 6 respectively).
274 Consistent with this, Likelihood Ratio Test (LRT) determined Model 0 to be the best model,
275 indicating that the extent of COVID-19 disease severity can be best explained by variability
276 in WHG ancestry fractions among European populations. Model 0 was found to be highly
277 significantly better than Model 1 (Temperature alone) and Model 2 (Humidity alone) ($P\approx 0$)
278 in recapitulating the variability of COVID-19 death/recovery ratio among the study
279 populations, suggesting that climatic factors such as colder temperature and variation in
280 humidity alone cannot explain the variability of COVID-19 disease in European countries.
281 Further, the combination of ancestry and temperature, ancestry and humidity (Model 3 and 4
282 respectively), and the combination of ancestry and climate (temperature + humidity) (Model
283 6) failed to explain variabilities in COVID-19 death/recovery ratio better than ancestry alone
284 (Model 0) ($P=0.88$, 0.36 and 0.61 respectively). This largely eliminates the possibility of
285 regional climate playing major role in the severity of COVID-19 presentation.

286 The association between WHG ancestry fraction (Model 0) and COVID-19
287 death/recovery ratio discernibly increased during reanalysis ($R^2=0.59$, $P=0.009$) (Table 2).
288 Consistent with the initial analysis, we did not find any association between 15 days' mean
289 temperature (Model 1) and humidity (Model 2) with COVID-19 death/recovery ratio (P
290 $=0.55$ and 0.31 respectively) and found significant association between the combination of
291 WHG ancestry and temperature (Model 3), WHG ancestry and humidity (Model 4), and
292 WHG ancestry and the combination of both climatic factors (temperature and humidity,
293 Model 6) with COVID-19 death/recovery ratio ($R^2=0.6$, 0.61 and 0.61 respectively) (Table
294 2). However, similar to our initial analysis, the significance of the test was driven by the
295 WHG ancestry fraction alone ($P=0.01$, 0.01 and 0.02 for Model 3, Model 4 and Model 6
296 respectively) and not by the climatic factors ($P_{\text{temperature}}=0.33$, $P_{\text{humidity}}=0.28$, and
297 $P_{\text{temperature}}=0.48$ and $P_{\text{humidity}}=0.35$ for Model 3, Model 4 and Model 6 respectively) (Table 2).
298 Finally, congruent with our initial analysis, LRT determined Model 0 to be the best model,
299 indicating COVID-19 manifestations can be best explained by variability in WHG ancestry
300 fractions among European populations.

301

302 **Table1:** West European hunter gatherers (WHGs) ancestry fractions, mean temperature (15
303 days), mean humidity (15 days), and COVID-19 death/recovery ratios of 10 European
304 countries employed in the study for initial analysis (data obtained for 5th April 2020)

Countries	Mean Temperature (°C, 15 days')	Mean Humidity (°C, 15 days')	WHG ancestry fraction (%)	Death/R recovery ratio 5 th April 2020	Linear Regression (R ² , P-value)
Norway	4.09	75.82	18.2	1.94	WHG ancestry only (<i>Model 0</i>) (R ² =0.34, P=0.03)
United Kingdom	8.18	69.45	26.1	31.94	Temperature only (<i>Model 1</i>) (R ² =2.402e-05, P=0.49)
France	12.64	74.09	20.9	0.49	Humidity only (<i>Model 2</i>) (R ² =0.04, P=0.29)
Spain	10	58.64	21.1	0.33	WHG ancestry + Temperature (<i>Model 3</i>) (R ² =0.34, P _{ancestry} =0.04, P _{temperature} =0.54)
Italy	10	62	23.1	0.73	WHG ancestry + Humidity (<i>Model 4</i>) (R ² =0.39, P _{ancestry} =0.04, P _{humidity} =0.23)
Greece	11.91	74.45	14.6	0.87	Temperature + Humidity (<i>Model 5</i>) (R ² =0.04, P _{temperature} =0.54, P _{humidity} =0.3)
Czech Republic	4.91	52.36	23.8	0.79	WHG ancestry + Climate (Temperature + Humidity) (<i>Model 6</i>) (R ² =0.4, P _{ancestry} =0.05, P _{temperature} =0.61, P _{humidity} =0.23)
Hungary	7.1	49.9	17.2	0.51	
Bulgaria	3.9	71	16.6	0.49	
Ukraine	6.8	45.4	17	1.28	

305

306

307 **Table 2:** West European hunter gatherers (WHGs) ancestry fractions, mean temperature (15

308 days), mean humidity (15 days), and COVID-19 death/recovery ratios of 10 European

309 countries employed in the study for reanalysis (data obtained for 30th June 2020)

Countries	Mean Temperature (°C, 15 days')	Mean Humidity (°C, 15 days')	WHG ancestry fraction (%)	Death/R recovery ratio 30 th June	Linear Regression (R ² , P-value)
-----------	---------------------------------	------------------------------	---------------------------	--	--

2020					
Turkey	22.9	58.3	4.2	0.03	WHG ancestry only (<i>Model 0</i>) ($R^2=0.59$, $P=0.009$)
Georgia	20.67	61.17	7.2	0.02	Temperature only (<i>Model 1</i>) ($R^2=0.002$, $P=0.55$)
France	19.27	68.82	20.9	0.42	Humidity only (<i>Model 2</i>) ($R^2=0.03$, $P=0.31$)
Spain	21.09	64.09	21.1	0.19	WHG ancestry + Temperature (<i>Model 3</i>) ($R^2=0.6$, $P_{\text{ancestry}}=0.007$, $P_{\text{temperature}}=0.33$)
Italy (TSI)	21	71.64	22.2	0.18	WHG ancestry + Humidity (<i>Model 4</i>) ($R^2=0.61$, $P_{\text{ancestry}}=0.007$, $P_{\text{humidity}}=0.28$)
Greece	23.1	58.8	14.6	0.14	Temperature + Humidity (<i>Model 5</i>) ($R^2=0.06$, $P_{\text{temperature}}=0.34$, $P_{\text{humidity}}=0.27$)
Armenia	18	70	9.8	0.03	WHG ancestry + Climate (Temperature + Humidity) (<i>Model 6</i>) ($R^2=0.61$, $P_{\text{ancestry}}=0.01$, $P_{\text{temperature}}=0.52$, $P_{\text{humidity}}=0.65$)
Hungary	19.7	72.5	17.2	0.22	
Bulgaria	16.6	76.2	16.6	0.08	
Romania	20	74.8	10.9	0.08	

310

311

312 Genome-wide association analyses (GWAS)

313

314 For GWAS, 753 genomes from across Europe with high COVID-19 death/recovery ratio

315 (case) were compared against 838 Eastern Asian genomes with relatively lower COVID-19

316 death/recovery ratio (control). Out of 597,573 SNPs employed in this study, 385,450 (64.5%)

317 markers revealed highly significant variation (Chi square ≥ 6.63 ; $P < 0.01$) between Europeans

318 and East Asians, indicating the discernible differences in the genetic tapestry of the two

319 populations (Fig. 2). For stringency, we annotated only the top 10,000 ranked SNPs

320 ($N=10,011$, Chi square ≥ 739 ; $P \leq 9.69 \times 10^{-163}$) using SNPnexus web-based server [30]

321 (Supplemental Table 2). Among top 10,000 ranked SNPs, 404 were associated with host

322 immune responsiveness including interferon (IFN) stimulated antiviral response, Interferon-
323 stimulated gene 15 (*ISG15*) mediated antiviral mechanism and 2'-5' oligoadenylate synthase
324 (OAS) mediated antiviral response. These 10,000 SNPs are linked to 161 genes associated
325 with immune system regulation including those related to innate immunity, antiviral response
326 and receptors (eg. *EDAR*) (Supplemental Table 3). Our results indicate that pathways such as
327 those involved in immunoregulatory interactions between a lymphoid and a non-lymphoid
328 cell, adaptive and innate immune system, IL-6 dependent immune responses, and
329 complement cascades show significant variation between genomes of East Asian and
330 European ancestry. Among these, type-I interferon and IL-6 dependent inflammatory
331 immune responses have been shown to be associated with SARS-CoV-2 infection [31].

332 Furthermore, the SNPs with the highest Chi square values (≥ 1500 ; $P \approx 0$) were found
333 to be associated with immune system related pathways such as Senescence-Associated
334 Secretory Phenotype (SASP), Interleukin-4 and Interleukin-13 signaling, TNFR2 non-
335 canonical NF- κ B pathway and Class I MHC mediated antigen processing and presentation
336 (Supplemental Table 4). Among these, Interleukin mediated 'cytokine storm' has been shown
337 to be associated with pulmonary inflammation among COVID-19 patients [32], and the
338 genetic variation in Class I MHC has been associated with severity of COVID-19
339 manifestations [9].

340 Finally, six immune system related SNPs (rs2243250, rs1800872, rs1800896,
341 rs1544410, rs1800629 and rs1805015) that display significant variability between Europeans
342 and East Asians (Odds Ratio > 2 , Chi square > 40 , $P < 10^{-10}$) are likely associated with the
343 development of immune responses in the first year of life contingent upon the gene-
344 environment interactions [33].

345

346 **Discussion**

347

348 The novel coronavirus SARS-CoV-2 infection has shown discernible worldwide variability.
349 While in some areas affected individuals are recovering relatively quicker, in others, recovery
350 times have been comparatively longer, and mortality high. Here, we sought to investigate the
351 likely association between genetic ancestry determinants and the extent of COVID-19
352 manifestations employing Europeans as a case study.

353 We found a significant positive correlation ($r=0.58$, $P=0.03$) between West European
354 hunter gatherers (WHG) ancestral fractions and COVID-19 death/recovery ratio which
355 discernibly increased ($r=0.77$, $P=0.009$) upon reanalyses as of 30th June 2020, removing

356 countries with small sample sizes and including those from intermediate geographical
357 locations between Europe and Asia, such as Georgia, Armenia and Turkey. We identified 404
358 immune response related SNPs by comparing publicly available data from 753 genomes
359 corresponding to various European countries against 838 genomes from various Eastern
360 Asian countries using GWAS. We found that SNPs associated with immune-system related
361 pathways such as interferon stimulated antiviral response, adaptive and innate immune
362 system and IL-6 dependent immune responses show significant differences in allele
363 frequencies [Chi square values (≥ 1500 ; $P \approx 0$)] between Europeans and East Asians.

364

365 The European story

366

367 Europe has been severely impacted owing to the ongoing COVID-19 pandemic. Our findings
368 revealed a significant positive correlation between WHG-related ancestry and COVID-19
369 death/recovery ratio ($P_{initial\ analysis}=0.03$, $P_{reanalysis}=0.009$). Further, we found significant linear
370 positive association between WHG ancestry fraction and COVID-19 death/recovery ratio
371 ($R^2_{initial\ analysis}=0.34$, $P_{initial\ analysis}=0.039$, $R^2_{reanalysis}=0.59$, $P_{reanalysis}=0.009$). Additionally, we
372 found that the regional climate alone is unlikely to play a major role in determining the
373 severity of COVID-19 disease. Previously it has been shown that European genomes evolved
374 uniquely with regards to immune responsiveness, particularly pertaining to responses against
375 viral infections [8]. Admixture with Neanderthals is believed to have introduced unique
376 regulatory variants into European genomes, which potentially regulated immune
377 responsiveness in European populations [7]. We surmise that the European genome
378 architecture has been extensively modulated by the complex origin and migration history of
379 modern-day Europeans during Paleolithic, Mesolithic and Neolithic periods that led to the
380 introduction of novel variants in European genomes.

381 Modern-day European genomic diversity has been thought to be shaped by variable
382 proportions of local hunter-gatherer ancestry (WHGs) [34]. While Neolithic, Iberian genomes
383 (modern day Spanish and Portuguese populations) revealed widespread evidence of WHG
384 admixture (10-27%), Neolithic German populations (Linearb and keramik– LBK culture)
385 revealed ~4–5% of the same [34]. Notably the variation in WHG ancestry fractions among
386 German and Spanish people correlated with the observed variabilities in COVID-19
387 death/recovery ratio in these two countries; German populations with discernably lower
388 WHG fractions have a current death/recovery ratio of 0.05, however, in Spain the ratio is
389 0.19, which corresponded to high WHG fraction among Spanish populations. A similar

390 correlation between WHG ancestry and the extent of COVID-19 manifestations was observed
391 for the UK, where GBR genomes depicted high WHG ancestry fractions (>25%) and was
392 correlated with high COVID-19 death/recovery ratio in the country (31.94). We note here
393 that due to unavailability of official COVID-19 recovery information from the UK since
394 April 2020, we were unable to include UK in our reanalyses.

395 Italy, one of the worst hit European countries during the COVID-19 pandemic
396 showed significant nationwide variation in the severity of disease manifestations
397 (death/recovery ratio=0.18) with a clear gradient in COVID-19 mortality rates from north to
398 south. While COVID-19 death/recovery ratio in Northern Italy is 0.19, the ratio is 0.15 in
399 Central Italy and 0.11 in Southern Italy. Interestingly, these numbers are consistent with
400 WHG fractions among Italian genomes. While Northern Italian genomes have ~23.1% WHG
401 ancestry fractions, the fraction reduces to 9.2% among Southern Italians (Fig. 1a and 1b).

402 In contrast, although COVID-19 cases have been rising rapidly in Russia, their
403 death/recovery ratio (0.02) has remained significantly lower than most European countries.
404 We assessed whether this is owing to their genomic make-up; all modern-day Russians
405 originated from two groups of East Slavic tribes: Northern and Southern [35] and have 0%
406 WHG ancestry fraction. They are genetically similar to modern-day other Slavic populations
407 such as Belarusians [35,36]. Consistent with this we observed significantly low COVID-19
408 death/recovery ratios in both Russia (0.02) and Belarus (0.01). In addition, COVID-19
409 manifestations were found to be far less detrimental in the Central Asian countries part of the
410 erstwhile Soviet Union (USSR) such as Kazakhstan (death/recovery ratio=0.01), Kyrgyzstan
411 (0.02), Tajikistan (0.01), and Uzbekistan (0.005).

412 Finally, Finnish genomes that were shaped predominantly by migrations from Siberia
413 ~3500 years ago [36] with very small WHG ancestral fractions, also have a discernibly
414 smaller COVID-19 death/recovery ratio (0.05) compared to neighboring Sweden, where the
415 ratio (0.88) is among the highest in world as of 31st May, 2020, after which no official
416 recovery information was available [2]. We note here that hunter gatherer genomes from
417 Sweden have shown to form a WHG cluster including Loschbour and La Braña2, indicating
418 high affinity of Swedish genomes towards WHG ancestry [37].

419 Historically, Turkey has been a crossroad of major population migrations across
420 Eurasia and served as a socio-cultural bridge between Europe and Asia. A whole genome-
421 based study revealed that while the modern-day Turkish population are genetically similar to
422 Southern European populations such as Italians, most Turkish genomes show signatures of
423 recent genetic contribution from ancestral East Asian populations [38]. On the other hand,

424 modern-day Armenians derived 29% of their European ancestry from Tyrolean Iceman [39],
425 a 5300-year-old Copper age individual, discovered along Tisenjoch Pass of the Ötztal Alps
426 [40]. Consequently, WHGs who contributed discernible ancestry fractions to most
427 Europeans, provided very low fraction (<10%) of the same in the ancestries of Near
428 Easterners (Turkey, Georgia and Armenia) [37]. Consistent with this, COVID-19
429 death/recovery ratio in Turkey, Georgia and Armenia remained discernibly lower compared
430 to most European countries (0.03, 0.02 and 0.03 respectively), further underscoring the
431 importance of WHG-related ancestral components in determining the severity of COVID-19
432 manifestations.

433 Overall, our results have revealed a clear association between WHG ancestry
434 fractions and the acuteness of COVID-19 manifestations suggesting the presence of unique
435 underlying genetic variants in European genomes that maybe correlated with increased
436 disease susceptibility and immune responsiveness. It is noteworthy that despite being better
437 equipped in terms of medical facilities and resources for testing and providing care to
438 COVID-19 affected individuals, most western and central European countries have struggled
439 substantially with significantly higher death/recovery ratios, as compared to many Asian
440 countries that are experiencing an increasing burden of patients with chronic underlying
441 conditions necessitating efficient health care systems with integrated care [41]. This further
442 illuminate that the intrinsic host genetic make-up as determined by ancestry is a major
443 determinant of the severity of COVID-19 disease.

444

445 Intrinsic 'protection' for East Asians?

446

447 As noted above, significant variation in allele frequencies of *ACE2* encoding for the host cell
448 receptor of SARS-CoV-2 has been reported between people of East Asian and European
449 ancestries [6]. The authors of this study speculated that the differences in allele frequencies of
450 *ACE2* coding variants were likely associated with variable expression of *ACE2* in tissues.

451 Additionally, they reported that most eQTL variants identified by them had discernibly
452 higher allele frequencies among East Asian populations compared to Europeans, which may
453 result in differential susceptibility towards SARS-CoV-2 infection in these populations under
454 similar conditions. Consistent with these findings we identified 404 SNPs that are associated
455 with host immune response, such as Interferon (IFN) stimulated antiviral response,
456 Interferon-stimulated gene 15 (ISG15) mediated antiviral mechanism and 2'-5' oligoadenylate
457 synthase (OAS) mediated antiviral response that showed large differences in allele

458 frequencies between Europeans and East Asians. The genetic differences between East and
459 Southeast Asians with Europeans can be largely attributed to their distinctive ancestral
460 origins. Most East Asians derive their ancestry from Mongolian hunter-gatherers who
461 dispersed over Northeast Asia 6000-8000 years ago [42]. Similarly, Southeast Asians exhibit
462 a mixture of East Asian ancestry (Southern Chinese agriculturalist) and a diverged form of
463 Eastern Eurasian hunter-gatherer ancestry (EHGs) [43]. People with similar ancestry can be
464 found as far south as Indonesia [43]. As noted above EHG ancestry fraction does not appear
465 to be significantly associated with COVID-19 manifestations. In congruence with their
466 unique ancestral make-up, all East and Southeast Asian countries exhibit discernibly low
467 COVID-19 death/recovery ratio compared to most west and central European countries
468 (Japan=0.06, South Korea=0.02, Hong Kong=0.006, Taiwan=0.01, Thailand=0.02,
469 Malaysia=0.01 and Singapore=0.0007). Notably, in mainland China, where the novel corona
470 virus originated, COVID-19 death/recovery ratio has remained discernibly low (0.06).
471 Overall, our results indicate that people of East and Southeast Asian ancestry appear to be
472 intrinsically protected against the most debilitating effects of SARS-CoV-2 infection.

473

474 India: a variegated canvas for COVID-19

475

476 Indians have a long and complex history of admixture between immigrant gene-pools
477 originating primarily in West Eurasia, Southeast Asia and the Indian hunter-gatherer lineage
478 with close genetic proximity to the present-day Andamanese people (Ancient Ancestral South
479 Indians: AASI) who likely arrived in India through the “southern exit” wave out of Africa
480 [44]. People of AASI ancestry admixed with an undivided ancient Iranian lineage that
481 subsequently split and lead to the formation of early Iranian farmers, herders, and hunter
482 gatherers approximately in the 3rd millennium BCE. This gene pool is referred to as the
483 ‘Indus Periphery’ gene pool [44,45], which is thought to be the major source of subsequent
484 peopling of India. Modern-day Indian genome is composed of largely four ancestral
485 components: Ancestral North Indian (ANI), Ancestral South Indian (ASI), Ancestral Tibeto-
486 Burman (ATB) and Ancestral Austro-Asiatic (AAA) [46]. ANI and ASI gene pools likely
487 arose around 2nd millennium BCE during the decline of Indus Valley Civilization (IVC),
488 which prompted multiple waves of migrations across the India. The southward migration of
489 Middle to Late Bronze Age people from Steppe (Steppe MLBA) into India is thought to have
490 coincided with the decline of IVC. It is speculated that the people of Indus-Periphery-related
491 ancestry, while migrating northward, admixed with Steppe MLBA immigrants to form the

492 ANI, while the others, who migrated southward and eastward admixed with AASI and
493 formed the ASI [45]. Austroasiatic speakers originated in Southeast Asia and subsequently
494 migrated to India during the Neolithic period. The admixture between local Indians and
495 incoming Southeast Asians took place ~2000-3800 years ago giving rise to the AAA
496 ancestry. Notably, the Indian population(s) with whom the incoming Southeast Asians
497 mingled were AASI related with little to no West Eurasian ancestry fraction [47]. Finally, it
498 has been showed that the Tibeto-Burmans (ATBs) derived their ancestry through admixture
499 with low-altitude East Asians who migrated from China and likely across Northern India or
500 Myanmar [48] leading to the high genomic proximity between Indians of ATB ancestry and
501 East Asians.

502 We believe that the severity of COVID-19 disease presentation is likely to vary
503 appreciably across the Indian populations. The genomic proximity between East Asians and
504 Indians mostly from Northeast India with prominent ATB ancestry, is reflected through lower
505 COVID-19 mortality rates from this region with Manipur, Mizoram, Nagaland and Sikkim
506 not registering any COVID-19 related death as of 30th June 2020 [19]. Notably the
507 death/recovery ratio in the largely ATB dominated region of Ladakh, India is 0.001,
508 indicating almost complete recovery so far amidst COVID-19 infected individuals from this
509 region.

510 Indian states with large number of indigenous tribal population with high prevalence
511 of AAA ancestry and/or with large fractions of AASI-related ancestry, eg. Andaman and
512 Nicobar Islands, Chhattisgarh, Jharkhand and Orissa have registered very few COVID related
513 deaths till date (death/recovery ratio 0, 0.005, 0.008, 0.004 respectively). Furthermore, Tamil
514 Nadu, despite ranking among the top three worst affected states in India, in terms of the
515 numbers of COVID-19 positive cases have registered significantly lower COVID-19
516 death/recovery ratio (0.02) that could be attributed to discernible fractions of AASI-related
517 ancestry among the people from this area. This indicates that Indians with predominant AAA
518 and AASI related ancestry may likely be less severely affected against the most detrimental
519 effects of COVID-19 infection. On the other hand, states with predominant ANI ancestry
520 such as Gujarat, Maharashtra, Madhya Pradesh and Delhi have shown discernibly higher
521 COVID-19 mortality rates (0.08, 0.08, 0.06 and 0.05, respectively).

522 We also note that approximately 83% of WHG ancestry has been associated with
523 mitochondrial DNA U haplogroup [49,50]. Overall, 13.1% Indians belong to U haplogroup
524 and are thus associated with WHG related ancestry [51]. And the U haplogroup frequency
525 shows discernible variation among Indian populations, while 23.3% North Indians belong to

526 U haplogroup, 10.3% South Indians belong to the same [51]. U haplogroup also shows
527 discernible variation across Indian societal structure, while 15% of Indian upper castes
528 belong to U haplogroup, 8% of Indian tribal populations (mostly of AAA ancestry) have the
529 same [52]. Taken together, overall ~8-19% Indians are likely to have WHG related ancestry.
530 Notably, our previous whole genome-based study found most Gujarati people to have ~3-4%
531 WHG related ancestry [53], similar to what is observed for Turkish Jews in the current study.
532 Our interrogation of worldwide populations suggests that the WHG ancestral fraction is
533 likely associated with acute COVID-19 manifestations and is predictive of debilitating effects
534 of COVID-19 infection among Indian populations with substantial fractions of WHG
535 ancestry. Overall COVID-19 manifestations in India is likely to be somewhat intermediate to
536 that observed in Europe and East Asia. This is also supported by the recent finding that allele
537 frequencies of *ACE2* variants among Indians are intermediate between East Asians and
538 Europeans [6].

539

540 Limitations and conclusion

541

542 The present study shines light of underlying genetic signatures that may be associated with
543 disparate COVID-19 severity and manifestations in worldwide populations. Nevertheless we
544 note that the current work has been performed using publicly available genomic data and a
545 more robust understanding in this regard will emanate from sequencing/genotyping
546 endeavours for COVID-19 patients across the spectrum of varied nationalities/ancestries and
547 geographical locations, including individuals with mild to moderate symptoms, severe
548 manifestations and death. We further note that since the current analyses is performed using
549 pre-existing genomic data, there is a disparity in number of individuals sequenced among
550 various populations which might influence the analyses in terms of statistical power and
551 errors.

552 We also note that populations from most European countries have higher mean ages
553 compared to India, which may accentuate mortality rates among Europeans. However, as
554 reported recently age alone may not suffice in exacerbating death/recovery ratios, and
555 underlying health conditions such as cardiovascular disease, chronic kidney disease,
556 hypertension, chronic respiratory disease, diabetes, cancers, HIV/AIDS, tuberculosis etc.,
557 many of which are already described as crucial comorbidity factors for COVID-19 may
558 modify disease manifestations and prognosis in patients [54]. It is imperative to consider here
559 that although the mean population age is lower in India compared to most European

560 countries, 7.7% Indians suffer from diabetes [55], 25.3% suffer from hypertension [56] and
561 approximately 14.5% suffer from some form of respiratory disorders linked to air pollution
562 [57], all of which are underlying health conditions that are reported to cause higher COVID-
563 19 death/recovery ratio [46]. Therefore, lower death/recovery ratio among Indians,
564 particularly in specific populations compared to most European countries highlights the role
565 of host genetic constitution in influencing COVID-19 disease progression.

566 Finally, several factors such as demographics (eg. population size and density),
567 healthcare facilities and hospital infrastructure, administrative preparedness and
568 implementation of social isolation and lockdowns were beyond the scope of the current study
569 but may influence the death/recovery metric of a country. Taken together our results strongly
570 advocate for the adoption of a rigorous worldwide population genetics driven approach to
571 expand and substantiate our current knowledge of SARS-CoV-2 infection, as well as
572 facilitate the development of population specific therapeutics to mitigate this worldwide
573 challenge.

574

575 **References**

576

- 577 1. Laboratory testing of 2019 novel coronavirus (2019-nCoV) in suspected human cases:
578 interim guidance, 17 January 2020 [Internet]. [cited 2020 Apr 9]. Available from:
579 <https://apps.who.int/iris/handle/10665/330676>
- 580 2. Dong E, Du H, Gardner L. An interactive web-based dashboard to track COVID-19 in
581 real time. *Lancet Infect Dis* 2020 20(5):533-534 PMID:32087114
- 582 3. Verity R, Okell LC, Dorigatti I, Winskill P, Whittaker C, Imai N, et al. Estimates of
583 the severity of coronavirus disease 2019: a model-based analysis. *Lancet Infect Dis*
584 2020 20(6):669-677. PMID: 32240634
- 585 4. Zhou P, Yang X-L, Wang X-G, Hu B, Zhang L, Zhang W, et al. A pneumonia
586 outbreak associated with a new coronavirus of probable bat origin. *Nature* 2020
587 Mar;579(7798):270-273. PMID:32015507
- 588 5. Lu R, Zhao X, Li J, Niu P, Yang B, Wu H, et al. Genomic characterisation and
589 epidemiology of 2019 novel coronavirus: implications for virus origins and receptor
590 binding. *The Lancet* 2020 Feb;395(10224):565–574. PMID: 32007145
- 591 6. Cao Y, Li L, Feng Z, Wan S, Huang P, Sun X, et al. Comparative genetic analysis of
592 the novel coronavirus (2019-nCoV/SARS-CoV-2) receptor ACE2 in different
593 populations. *Cell Discov* 2020 Dec;6(1):11. PMID: 32133153

- 594 7. Stawiski EW, Diwanji D, Suryamohan K, Gupta R, Fellouse FA, Sathirapongsasuti F,
595 Liu J, Jiang YP, Ratan A, Mis M, Santhosh D. Human ACE2 receptor polymorphisms
596 predict SARS-CoV-2 susceptibility. bioRxiv. 2020 Jan 1.
- 597 8. Li G, Fan Y, Lai Y, Han T, Li Z, Zhou P, et al. Coronavirus infections and immune
598 responses. *J Med Virol* 2020 Apr;92(4):424–432. PMID: 31981224
- 599 9. Nguyen A, David JK, Maden SK, Wood MA, Weeder BR, Nellore A, Thompson RF.
600 Human leukocyte antigen susceptibility map for SARS-CoV-2. *Journal of virology*.
601 2020 Jun 16;94(13):e00510-20. PMID: 32303592.
- 602 10. Kuo CL, Pilling LC, Atkins JL, Masoli JA, Delgado J, Kuchel GA, et al. APOE e4
603 genotype predicts severe COVID-19 in the UK Biobank community cohort. *J*
604 *Gerontol A Biol Sci Med Sci* 2020 May 26;glaa131. PMID: 32451547
- 605 11. Quach H, Rotival M, Pothlichet J, Loh Y-HE, Dannemann M, Zidane N, et al.
606 Genetic Adaptation and Neandertal Admixture Shaped the Immune System of Human
607 Populations. *Cell* 2016 Oct 20;167(3):643-656.e17. PMID:27768888
- 608 12. Mathieson I, Alpaslan-Roodenberg S, Posth C, Szécsényi-Nagy A, Rohland N,
609 Mallick S, et al. The genomic history of southeastern Europe. *Nature* 2018
610 Mar;555(7695):197–203. PMID: 29466330
- 611 13. Olalde I, Brace S, Allentoft ME, Armit I, Kristiansen K, Booth T, et al. The Beaker
612 phenomenon and the genomic transformation of northwest Europe. *Nature* 2018
613 Mar;555(7695):190–196. PMID: 29466337
- 614 14. Olalde I, Mallick S, Patterson N, Rohland N, Villalba-Mouco V, Silva M, et al. The
615 genomic history of the Iberian Peninsula over the past 8000 years. *Science* 2019 Mar
616 15;363(6432):1230–1234. PMID: 30872528
- 617 15. Kachuri L, Francis SS, Morrison M, Bossé Y, Cavazos TB, Rashkin SR, et al. The
618 landscape of host genetic factors involved in infection to common viruses and SARS-
619 CoV-2. medRxiv. 2020 Jan 1. PMID: 32511533
- 620 16. David Reich Lab datasets URL <https://reich.hms.harvard.edu/datasets> [accessed 2020-
621 03-25]
- 622 17. Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D. Principal
623 components analysis corrects for stratification in genome-wide association studies.
624 *Nat Genet* 2006 Aug;38(8):904–909. PMID: 16862161
- 625 18. Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, et al.
626 PLINK: A Tool Set for Whole-Genome Association and Population-Based Linkage
627 Analyses. *Am J Hum Genet* 2007 Sep;81(3):559–575. PMID: 17701901

- 628 19. India COVID-19 statewise status URL [https://www.mygov.in/corona-data/covid19-](https://www.mygov.in/corona-data/covid19-statewise-status)
629 [statewise-status](https://www.mygov.in/corona-data/covid19-statewise-status)
- 630 20. Haak W, Lazaridis I, Patterson N, Rohland N, Mallick S, Llamas B, et al. Massive
631 migration from the steppe was a source for Indo-European languages in Europe.
632 Nature 2015 Jun;522(7555):207–211. PMID: 25731166
- 633 21. Patterson N, Moorjani P, Luo Y, Mallick S, Rohland N, Zhan Y, et al. Ancient
634 Admixture in Human History. Genetics 2012 Nov;192(3):1065–1093. PMID:
635 22960212
- 636 22. Alexander DH, Novembre J, Lange K. Fast model-based estimation of ancestry in
637 unrelated individuals. Genome Res 2009 Sep 1;19(9):1655–1664. PMID: 19648217
- 638 23. Lazaridis I, Nadel D, Rollefson G, Merrett DC, Rohland N, Mallick S, et al. Genomic
639 insights into the origin of farming in the ancient Near East. Nature. 2016
640 Aug;536(7617):419-24.
- 641 24. Graphpad Prism 8 URL www.graphpad.com
- 642 25. Time and Date website URL <https://www.timeanddate.com/>
- 643 26. World Meteorological Organization URL <https://public.wmo.int/en>
- 644 27. Meteorological Assimilation Data Ingest System (MADIS) URL
645 <https://madis.ncep.noaa.gov/>
- 646 28. Team QD. 2016. QGIS geographic information system. Open Source Geospatial
647 Foundation Project. URL <http://qgis.osgeo.org>
- 648 29. Barrett JC, Fry B, Maller J, Daly MJ. Haploview: analysis and visualization of LD
649 and haplotype maps. Bioinformatics 2005 Jan 15;21(2):263–265. PMID: 15297300
- 650 30. Dayem Ullah AZ, Oscanoa J, Wang J, Nagano A, Lemoine NR, Chelala C.
651 SNPnexus: assessing the functional relevance of genetic variation to facilitate the
652 promise of precision medicine. Nucleic Acids Res 2018 Jul 2;46(W1):W109–W113.
653 PMID: 29757393
- 654 31. Ramlall V, Thangaraj PM, Tatonetti NP, Shapira SD. Identification of Immune
655 complement function as a determinant of adverse SARS-CoV-2 infection outcome.
656 medRxiv. 2020 Jan 1. PMID: 32511494
- 657 32. Coperchini F, Chiovato L, Croce L, Magri F, Rotondi M. The cytokine storm in
658 COVID-19: an overview of the involvement of the chemokine/chemokine-receptor
659 system. Cytokine & Growth Factor Reviews. 2020 May 11. PMID: 32446778
- 660 33. SNPedia URL <https://www.snpedia.com>

- 661 34. Lipson M, Szécsényi-Nagy A, Mallick S, Pósa A, Stégnár B, Keerl V, et al. Parallel
662 palaeogenomic transects reveal complex genetic history of early European farmers.
663 Nature 2017 Nov;551(7680):368–372. PMID: 29144465
- 664 35. Morozova I, Evsyukov A, Kon'kov A, Grosheva A, Zhukova O, Rychkov S. Russian
665 ethnic history inferred from mitochondrial DNA diversity. Am J Phys Anthropol 2012
666 Mar;147(3):341-51. PMID: 22183855
- 667 36. Lamnidis TC, Majander K, Jeong C, Salmela E, Wessman A, Moiseyev V, et al.
668 Ancient Fennoscandian genomes reveal origin and spread of Siberian ancestry in
669 Europe. Nat commun. 2018 Nov 27;9(1):5018. PMID:30479341
- 670 37. Lazaridis I, Patterson N, Mittnik A, Renaud G, Mallick S, Kirsanow K, Sudmant PH,
671 Schraiber JG, Castellano S, Lipson M, Berger B. Ancient human genomes suggest
672 three ancestral populations for present-day Europeans. Nature. 2014
673 Sep;513(7518):409-13. PMID:25230663
- 674 38. Alkan C, Kavak P, Somel M, Gokcumen O, Ugurlu S, Saygi C, et al. Whole genome
675 sequencing of Turkish genomes reveals functional private alleles and impact of
676 genetic interactions with Europe, Asia and Africa. BMC genomics. 2014 Dec
677 1;15(1):963. PMID:25379065
- 678 39. Haber M, Mezzavilla M, Xue Y, Comas D, Gasparini P, Zalloua P, et al. Genetic
679 evidence for an origin of the Armenians from Bronze Age mixing of multiple
680 populations. European Journal of Human Genetics. 2016 Jun;24(6):931-6.
681 PMID:26486470
- 682 40. Keller A, Graefen A, Ball M, Matzas M, Boisguerin V, Maixner F, Leidinger P, et al.
683 New insights into the Tyrolean Iceman's origin and phenotype as inferred by whole-
684 genome sequencing. Nat commun. 2012 Feb 28;3(1):1-9. PMID:22426219
- 685 41. Tham TY, Tran TL, Prueksaritanond S, Isidro JS, Setia S, Welluppillai V. Integrated
686 health care systems in Asia: an urgent necessity. Clinical interventions in aging.
687 2018;13:2527. PMID:30587945
- 688 42. Wang C-C, Yeh H-Y, Popov AN, Zhang H-Q, Matsumura H, Sirak K, et al. The
689 Genomic Formation of Human Populations in East Asia [Internet]. Genomics; 2020
690 Mar. [doi: 10.1101/2020.03.25.004606]
- 691 43. Lipson M, Cheronet O, Mallick S, Rohland N, Oxenham M, Pietrusewsky M, et al.
692 Ancient genomes document multiple waves of migration in Southeast Asian
693 prehistory. Science 2018 Jul 6;361(6397):92–95. PMID: 29773666

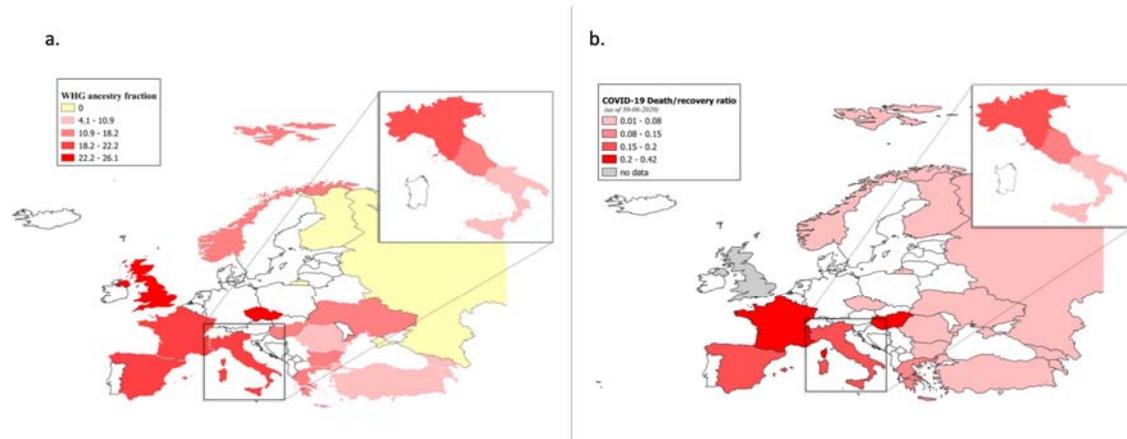
- 694 44. Shinde V, Narasimhan VM, Rohland N, Mallick S, Mah M, Lipson M, et al. An
695 Ancient Harappan Genome Lacks Ancestry from Steppe Pastoralists or Iranian
696 Farmers. *Cell* 2019 Oct;179(3):729-735.e10. PMID: 31495572
- 697 45. Narasimhan VM, Patterson N, Moorjani P, Rohland N, Bernardos R, Mallick S, et al.
698 The formation of human populations in South and Central Asia. *Science* 2019 Sep
699 6;365(6457):eaat7487. PMID: 31488661
- 700 46. Basu A, Sarkar-Roy N, Majumder PP. Genomic reconstruction of the history of extant
701 populations of India reveals five distinct ancestral components and a complex
702 structure. *Proc Natl Acad Sci* 2016 Feb 9;113(6):1594–1599. PMID: 26811443
- 703 47. Tätte K, Pagani L, Pathak AK, Köks S, Ho Duy B, Ho XD, et al. The genetic legacy
704 of continental scale admixture in Indian Austroasiatic speakers. *Sci Rep* 2019
705 Apr;9(1):6104. PMID:30967570
- 706 48. Gneccchi-Ruscione GA, Jeong C, De Fanti S, Sarno S, Trancucci M, Gentilini D, et al.
707 The genomic landscape of Nepalese Tibeto-Burmans reveals new insights into the
708 recent peopling of Southern Himalayas. *Sci Rep* 2017 Dec;7(1):15512. PMID:
709 29138459
- 710 49. Molto JE, Loreille O, Mallott EK, Malhi RS, Fast S, Daniels-Higginbotham J, et al.
711 Complete mitochondrial genome sequencing of a burial from a romano–christian
712 cemetery in the Dakhleh Oasis, Egypt: preliminary indications. *Genes*. 2017
713 Oct;8(10):262. PMID:28984839
- 714 50. Fu Q, Rudan P, Pääbo S, Krause J. Complete mitochondrial genomes reveal Neolithic
715 expansion into Europe. *PloS one*. 2012 Mar 13;7(3):e32473. PMID:22427842
- 716 51. Kivisild T, Bamshad MJ, Kaldma K, Metspalu M, Metspalu E, Reidla M, et al. Deep
717 common ancestry of Indian and western-Eurasian mitochondrial DNA lineages.
718 *Current Biology*. 1999 Nov 18;9(22):1331-4. PMID:10574762
- 719 52. Karmin M. Human mitochondrial DNA haplogroup R in India: Dissecting the
720 phylogenetic tree of South Asian-specific lineages [Master’s thesis]. University of
721 Tartu; 2005.
- 722 53. Das R, Upadhyai P. Investigating the west Eurasian ancestry of Pakistani Hazaras.
723 *Journal of genetics*. 2019 Jun 1;98(2):43. PMID:31204712
- 724 54. Clark A, Jit M, Warren-Gash C, Guthrie B, Wang HH, Mercer SW, et al. Global,
725 regional, and national estimates of the population at increased risk of severe COVID-
726 19 due to underlying health conditions in 2020: a modelling study. *Lancet Glob*
727 *Health*. 2020 Jun 15. PMID:32553130

728 55. Tandon N, Anjana RM, Mohan V, Kaur T, Afshin A, Ong K, et al. The increasing
729 burden of diabetes and variations among the states of India: the Global Burden of
730 Disease Study 1990–2016. *The Lancet Global Health*. 2018 Dec 1;6(12):e1352-62.
731 PMID:30219315

732 56. Jose AP, Prabhakaran D. World Hypertension Day: Contemporary issues faced in
733 India. *The Indian journal of medical research*. 2019 May;149(5):567-570.
734 PMID:31417023

735 57. Singh V, Sharma BB. Respiratory disease burden in India: Indian chest society
736 SWORD survey. *Lung India: Official Organ of Indian Chest Society*. 2018
737 Nov;35(6):459-460. PMID:30381552

738
739
740
741
742
743
744
745
746
747
748
749
750
751
752
753
754
755
756
757
758
759
760
761
762
763



764

765 Fig. 1: Maps showing WHG ancestry fractions (a) and COVID-19 death/recovery ratio as of
766 30th June 2020 (b) in various European countries. Maps were plotted in QGIS software
767 (<http://qgis.osgeo.org>). To show gradient decrease in WHG ancestry fractions, WHG ancestry
768 proportion in Central Italy (a: inset) was calculated by taking an average of WHG ancestry
769 proportions of Northern (23.1%) and Southern (9.2%) Italy. Finland and Russia which could
770 not be modelled to WHG ancestry are shown in yellow (a) and since current COVID-19
771 recovery rate in UK is unknown, it is shown in grey (b).

772

773

774

775

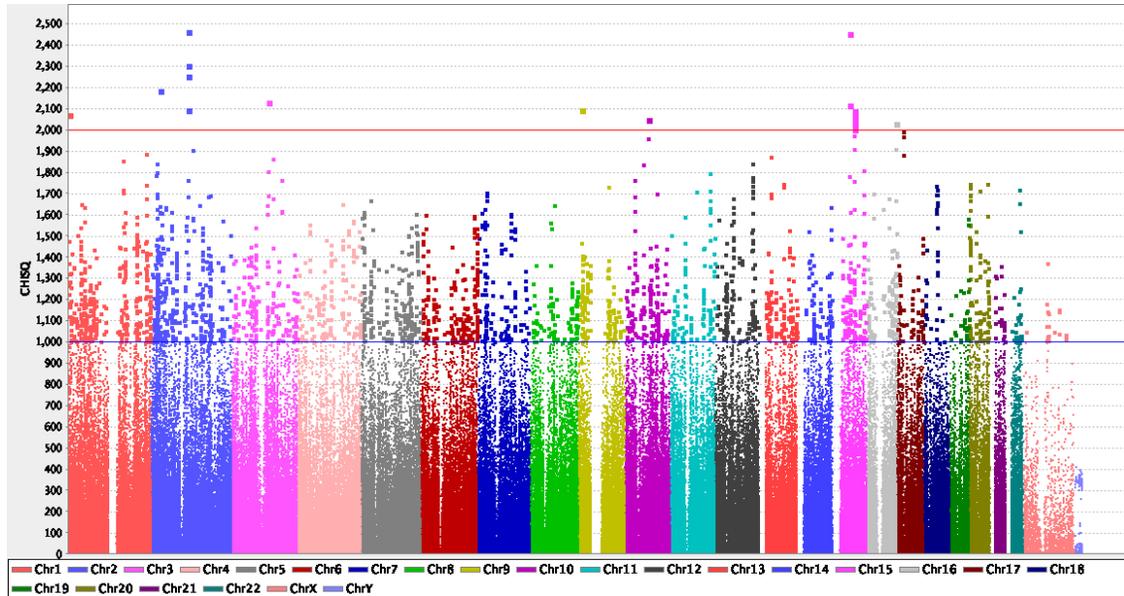
776

777

778

779

780



781

782

783 Fig.2: Manhattan Plot. X-axis represents chromosomes (chr 1 to chr Y). SNPs present in the
784 chromosomes are designated with dots. The chi square values are plotted in Y-axis. GWAS
785 with 753 European genomes were compared to 838 Eastern Asian genomes. Out of 597,573
786 SNPs employed, 385,450 (64.5%) markers revealed highly significant variation between
787 Europeans and East Asians. The SNPs with chi square values >1000 are indicated with the
788 blue line and those with chi square value >2000 are indicated with the red line. While 18
789 SNPs showed chi square values >2000, 2,992 SNPs had chi square values >1000.

790

791

792

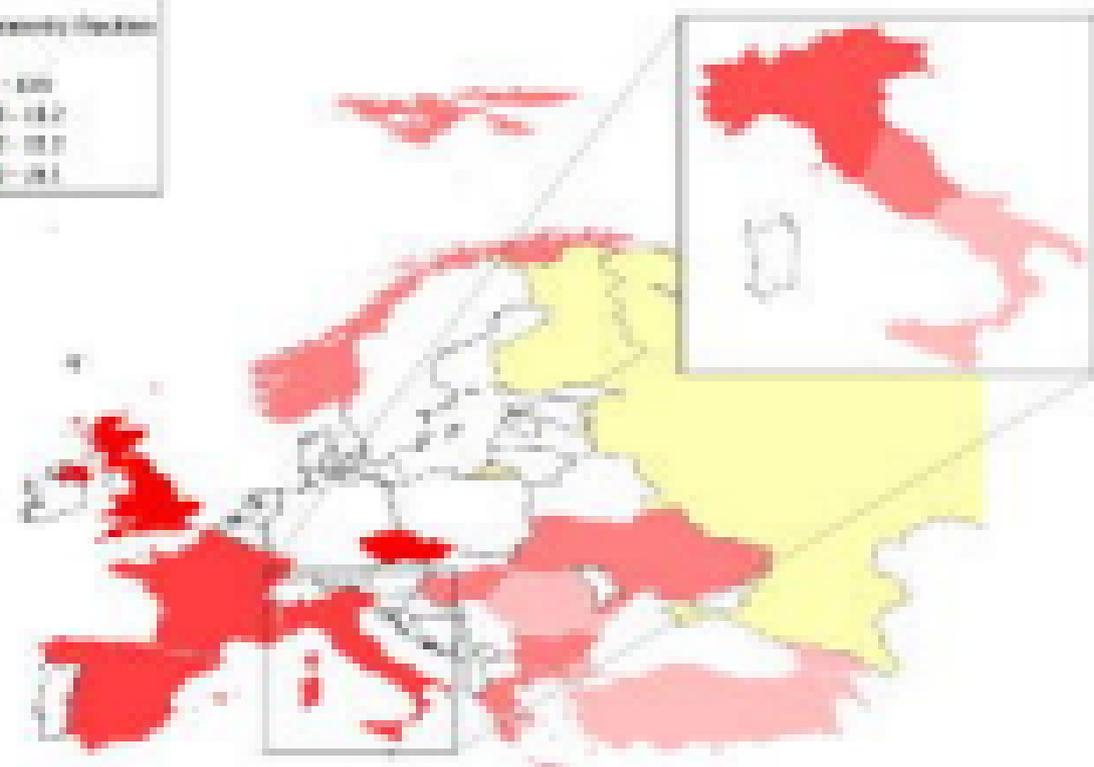
793

794

795

796

a.



b.

