Genetic diversity, disease prevalence and the coronavirus pandemic

Phiri, Andrew

Department of Economics, Nelson Mandela University

17 June 2020

GENETIC DIVERSITY, DISEASE PREVELANCE AND THE CORONAVIRUS PANDEMIC

Andrew Phiri

Department of Economics, Faculty of Business and Economic Studies, Nelson Mandela University, Port Elizabeth, South Africa, 6031

Abstract: The COVID-19 disease outbreak is the deadliest viral pandemic our generation has experienced, and much uncertainty exists over the vulnerability of different populations to the virus since a clinically-approved vaccination does not exist. Our study investigates whether evolutionary processes such as genetic diversity and cultural behaviour norms can explain the differences in COVID-19 virus infections and mortalities observed in different countries. Using a sample of 133 countries we find that populations with higher expected genetic heterozygosity and more historical exposure to infectious diseases are associated with lower COVID-19 infections and mortalities. Further investigations reveal two 'channels' of transmission. Firstly, a longer migratory distance from the origins of homo sapiens adversely influences expected heterozygosity, which then increases the populations susceptibility to the COVID_19 virus. Secondly, higher disease prevalence leads to higher collectivism (lower individualism) behaviour, which then reduces the populations susceptibility to COVID_19 infections. Our analysis is robust to the inclusion of additional controls and dummies. Policy implications of our findings are discussed.

Keywords: COVID-19; Expected heterozygosity; Disease prevalence; Collectivism; Individualism; Deep roots.

JEL classification: Expected heterozygosity; Historical prevalence to infectious disease (HPPI); Individualism/Collectivism; Coronavirus.

1. INTRODUCTION

"*...humans who domesticated animals [fell] victim to the newly developed germs, but those humans evolved substantial resistance to new[er] disease...*" (Diamond, 1997:92).

The above quote is taken from Jared Diamonds Non-fictional Pulitzer Winning Book '*Guns, Germs and Steel: The fates of human societies*' which gives a remarkable archaeological account of human development spanning from the end of the Ice Age over 13,000 years ago up to the New World. In the 11th Chapter of the book, Diamond (1997) focuses on the emergence and development of viral diseases which became prominent around the Neolithic transition when human society traversed from hunter-gathering livelihoods towards agricultural based societies. At the centre of this revolution is the domestication of different animals (mainly for food, clothing, farming and travel) which were adaptable to both the climate and illnesses of these agricultural societies but brought about infectious diseases confined exclusively to human beings. These viral diseases, which are concomitant to human evolution, are branded as "*...the biggest killers of people...*" (Diamond, 1997: 197) even when compared to fatalities resulting from Wars. The two deadliest viral pandemics experienced over the last millennium are the Black Death of 1346-1352 which claimed the lives of more than two-fifths of Europe's population (Jedwab et al., 2019) as well as the Spanish influenza of 1918-1921 which was responsible for between 50-100 million deaths globally (i.e. approximately 5-10% of the then World's population) (Karlsson et al., 2014).

The World is currently afflicted by the coronavirus (COVID-19) epidemic which is labelled as the deadliest outbreak of viral disease since the 'Spanish flu' a Century ago. According to Phan (2020), 'Patient Zero' of the COVID-19 disease was identified in December 2019 in Wuhan, China and due to the rapid spread of the disease worldwide, the World Health Organization (WHO) officially declared it a global pandemic in March 11th when global infections had reached 126,214 and total deaths recorded at 4,628. Despite the World having built stronger health and research institutions in comparison to those which existed during the time of the Spanish flu, there is however one predicament; no vaccine or cure for the disease exists at this moment. The best option for governments worldwide is to focus on flattening the 'epidemic curve' through quarantining infected person, encouraging social distancing, placing traveling restrictions and implementing other emergency 'lockdown' strategies even though the literature casts much ambiguity surrounding the effectiveness of these strategies. In a quasi-

experiment performed for France, Adda (2016) uses high frequency data on 'school closures' and 'public transportation' to demonstrate on how shutdown policies may be successful in curbing the viral spread of flu-like illnesses and yet this comes at a cost of excessive productivity losses. Adda (2016) concludes that lockdown strategies are most cost-effective only when the actual death rate is above it's average. Xiao and Torok (2020) further caution that the prolonged closure of schools and business, particularly in less-industrialized countries, could result in civil unrest which will reduce societies compliance with lockdown measures. Sibony (2020) refers to this phenomenon as 'behavioural fatigue' and cites this as factor which could offset a society's fear of the pathogen hence leading to more risky behaviour. More recently, Toda (2020) calibrates a Susceptible-Infected-Recovered (SIR) epidemic model using global COVID-19 data to highlight on the ineffectiveness of 'draconian' lockdown strategies used by governments in attempts to lower transmission rate of the disease. Toda's (2020) calibrations reveal that for governments to significantly reduce the spread of the disease they would have to employ strict lockdown measures for periods exceeding 12 weeks which the author finds to be economically unsustainable. Moreover, the findings show that current measures taken to lower the transmission rate have instead lowered the chances of populations acquiring 'herd immunity' and this increases the risk of the epidemic resurfacing in future periods.

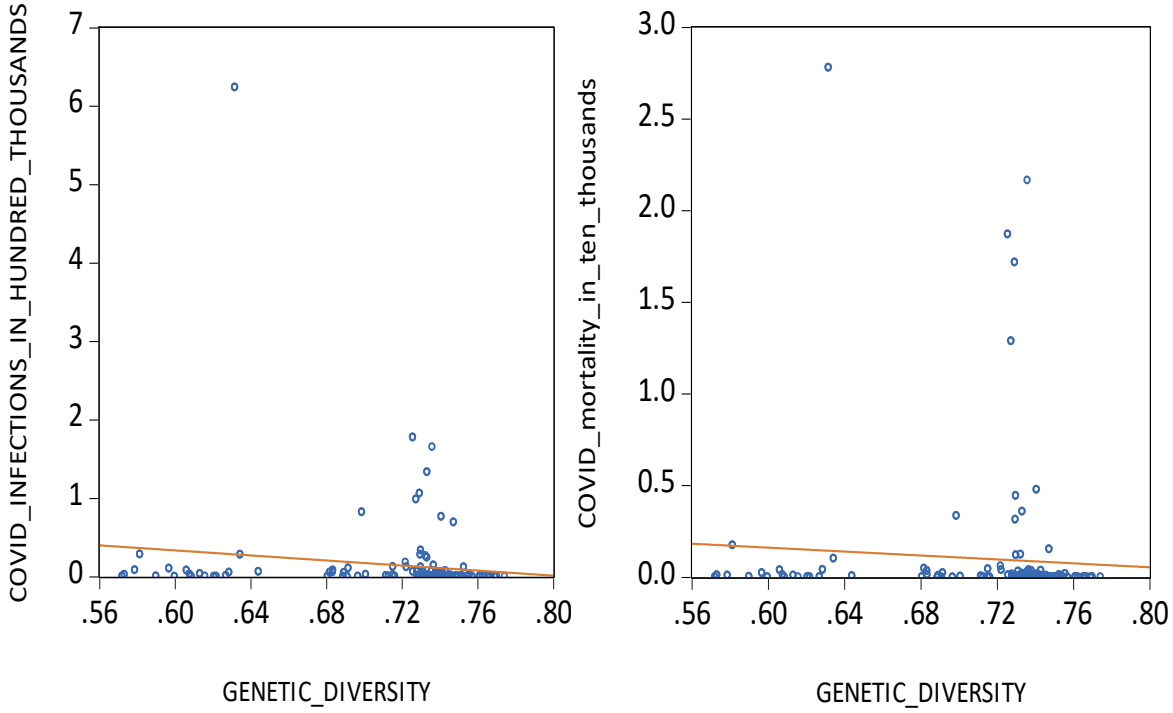
Notably, the COVID-19 morbidity and mortality levels vastly differ amongst many countries across the globe, with 16 countries even reporting zero cases at the time of writing (i.e. Comoros, Kiribati, Lesotho, Marshall Islands, Micronesia, Nauru, North Korea, Palau, Samoa, Sao and Principles, Solomon Island, Tajikistan, Tonga, Turkmenistan, Tuvalu and Vanuatu). What is even more striking is that poorer countries characterized by less sophisticated health systems, particularly in the Sub-Saharan Africa (SSA), South Asia and the Oceania Islands regions, exhibit fewer infections and death cases compared to those reported in more industrialised economies. Given the uncertainty of a vaccine being formulated against the COVID-19 disease, the varying levels of morbidities and mortalities experienced in different countries worldwide are more likely to be determined by other factors unrelated to the quality of health and medical institutions. So far, scientists have established that the most vulnerable groups towards COVID-19 infection and mortality include elderly populations (Koff and Wlliams, 2020), persons with ill health and comorbidities (Yang et al., 2020), males as opposed to females (Wenham et al. (2020) and in populations whose infants were not previously vaccinated with Bacillus Calmette-Guerin (BCG) (Miller et al., 2020). Our study

goes beyond these demographic factors and proposes evolutionary factors like genetic heterozygosity and historical prevalence to infectious disease, as deeper explanations of susceptibility to the pandemic and to reach our objective we borrow from two scientific disciplines of research.

Firstly, we draw from mainstream genetic theory which hypothesizes on genetically 'homogenous' populations being more susceptible to viral infection and disease progression compared to more genetically diverse populations (King and Lively, 2012; Anacleto et al., 2019). One of the most compelling proofs of this proposition is presented by Lively (2010) who uses a mathematical epidemiological model to demonstrate an inverse relationship between the average intrinsic rate of viral infections and the number of host genotypes in a population. We also draw from the 'Out of Africa' hypothesis modelled by a separate group of population geneticists which further predicts on a natural selection evolutionary process in which populations only carried a sub-set of genetic material when they migrated away from their parental colonies whose origins trace to a common ancestor in Addis Ababa (Prugnolle et al. (2005); Deshpande et al. (2008)). To empirically test the hypothesis, Ramachandran et al. (2005) and Ashraf and Galor (2013) use the allelic frequencies of 377 loci to measure the changes in genetic heterozygosity for 51 ethnic groups relating to different population settlements along the 5 migratory paths leading 'Out of Africa'. The authors find that expected heterozygosity at these microsatellite loci not only decrease along the migratory distance from East Africa but also the genetic variation between populations in various settlements outside Africa is larger when there is a longer migratory distance between the populations i.e. serial-founder effect. Notably, these measures of expected heterozygosity have been empirically used by Unified Growth Theorists (UGT) to explain global differences in human capital development (Sequeira et al., 2019), technological advancements (Sequeira and Santos, 2019) and susceptibility to conflict (Arbatli et al., 2020). Our study uses Ashraf and Galor's (2013) measures of expected heterozygosity to empirically examine whether Lively's (2010) hypothesis of an inverse relationship between genetic diversity and the spread of viral infection holds for the case of the ongoing COVID-19 pandemic. Figure 1 presents a scatterplot between genetic heterozygosity for 133 countries and their corresponding COVID-19 infections/mortalities at 15st April 2020. A preliminary fit of the data reveals an inverse co-relationship between genetic diversity and the COVID-19 virus which provides the basis for our study's first testable hypothesis i.e.

H_1: Genetic diversity is inversely related with COVID-19 infections.

Figure 1: COVID infections, mortalities and genetic diversity

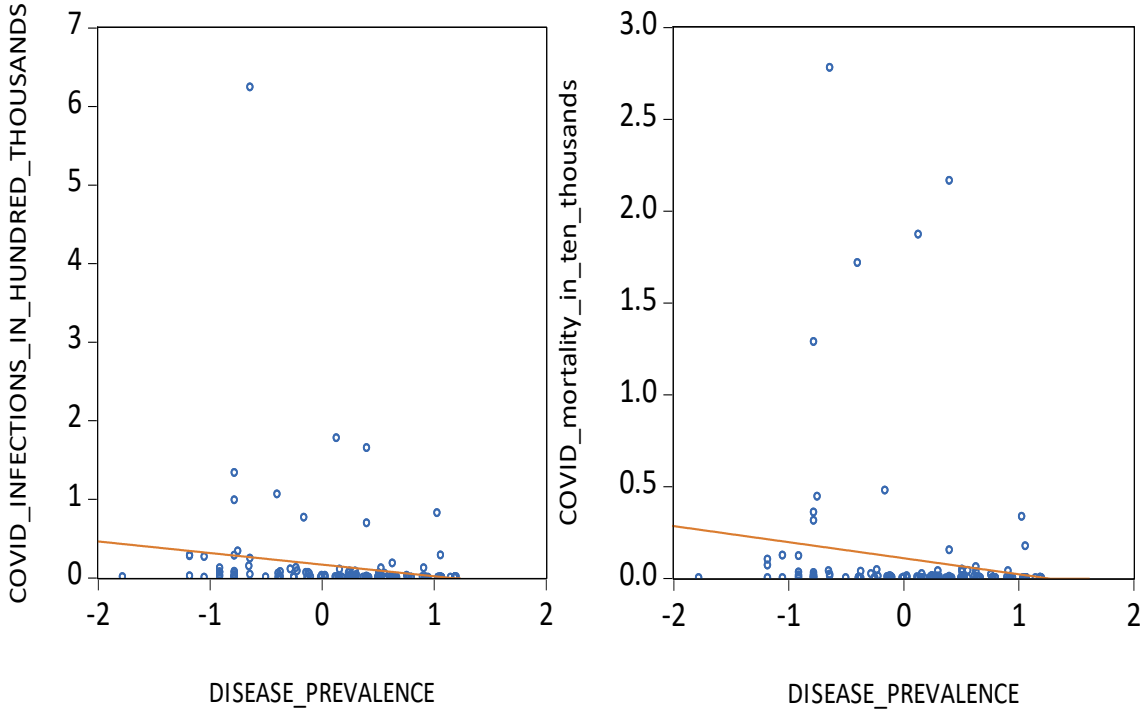


Secondly, our study draws from the sociocultural literature and we particularly focus on the 'pathogenic stress theory' which explains the behavioural influence that parasite and infection stress has on the development of cultural norms and individual values (Fincher et al. (2008), Fincher and Thornhill (2008, 2012), Murray and Schaller (2010), Murray et al. (2008, 2011)). According to the theory, the threat of diseases on the survival of ancestral populations led to natural selection pressures whereby societies with higher exposure to previous pathogenic infections evolved psychological and behavioural 'antipathogen' defences against novel diseases. Adopted cultural traits such as the use of spices as a natural antibiotic in the preparation of food (Billing and Sherman, 1998), limited interaction with 'out-group' members (Fincher et al., 2008), mate preferences (Murray et al., 2011) and prejudice against people perceived as unhealthy, unclean or unhygienic (Fincher and Thornhill, 2012) are believed to reflect 'collectivism' behaviour which helps these societies avoid the infection and spread of newer diseases. Murray et al (2008) and Murray and Schaller (2010) use epidemiological atlases to code an index of z-sores which capture the prevalence of 9 different types of pathogens causing infectious diseases (i.e. leishmanias, schistosomes, trypanosomes, leprosy, malaria, typhus, filariae, dengue and tuberculosis) and estimate positive (negative) correlations

with measures of collectivism (individualism) presented in Hofstede (2001), Suh et al. (1998), Gelfand et al. (2004) and Kashima and Kashima (1998). Similar findings are observed in Cashdan and Steele (2013), Nikolaev and Salahodjaev (2017) and Ang (2019) albeit using different regression control variables in their respective analysis. However, these previous studies have not examined whether societies with higher disease prevalence are less susceptibility to novel viral infection, as hypothesized by the 'pathogenic stress theory'. Our paper uses the measures of disease prevalence constructed by Murray and Schaller (2010) to investigate the empirical relationship between pathogen prevalence and the ongoing COVID-19 virus. To further motivate our study, we present a preliminary scatterplot between the HPPI index for 133 countries and their corresponding COVID-19 infections/mortalities in Figure 2. A preliminary fit of the data provides visual support for a negative correlation between the variables and this leads to the formation of a second testable hypothesis i.e.

H_2: Disease prevalence is inversely related with COVID-19 infections.

Figure 2: COVID infections, mortalities and disease prevalence



All-in-all, our study enriches the knowledge on susceptibility of populations to the coronavirus disease in three ways. Firstly, whilst most previous studies focus on demographic factors limited to regional data, our paper presents an analysis using global data covering 133

countries worldwide which makes our study more relevant for decision making at a global level. Secondly, we provide evidence on evolutionary process and behaviour traits as being significantly correlated with coronavirus cases even after controlling for a host of plausible geographical, climatic and ecological determinants of general disease. We find that the magnitude of effect for expected heterozygosity and disease prevalence on the COVID-19 virus is greater and more significant in comparison to other control factors. These findings have important implications for biomedical and pharmaceutical research in their quest towards formulating a vaccine as well as for behavioural policies used by governments to control the spread of the COVID-19 virus in the absence of a cure for the disease. Lastly, we address possible endogeneity problems by making use of instrument variable methodology. On one hand, we follow Ashraf and Galor (2013) and use migratory distance from origin as an instrument for expected heterozygosity, which we then use as a predictor for COVID-19 cases. On the other hand, we follow Nikolaev and Salahodjaev (2017) and Ang (2019) and use disease prevalence as an instrument for collectivism/individualism, which we then use as a predictor for COVID-19 cases. The findings obtained from the instrumental variable estimates support the channels of causality implied in Diamond's (1997) work.

We proceed with the rest of the study as follows. The next section outlines the empirical regressions used in our study and describes data to be used in our empirical analysis. Section 3 presents the baseline OLS empirical estimates of the regressions whilst section 4 presents sensitivity analysis. Section 5 presents the two-staged least squares (2SLS) estimates to address possible endogeneity in the regressions. Our paper is concluded in Section 6 in the form implications for policymakers and researchers.

2. METHODS AND DATA

Consistent with the two hypotheses specified in the introduction of the study, we model two cross-country least squares regressions for estimation purposes. Firstly, we model coronavirus infections/mortalities (SARS-CoV-2) as being endogenous to genetic diversity (GEN_DIV) and other conditioning variables (CONTROLS) i.e.

$$SARS\text{-}CoV\text{-}2_i = \mu + \beta\ GEN\_DIV_i + CONTROLS_i + \varepsilon_i \quad (1)$$

Secondly, we model coronavirus infections/mortalities (SARS-CoV-2) as being endogenous to pathogen prevalence (HPPI) and other control variables i.e.

$$SARS\text{-}CoV\text{-}2_i = \mu + \beta\ HPPI + CONTROLS_i + e_{t,\,i} \qquad (2)$$

The dependent variable in equations (1) and (2) is measured by total morbidities and total mortalities of SARS-CoV-2 virus in 133 developing and developed economies sourced from the John Hopkins database and consists of total infections and mortalities at 28/04/2020. The list of countries used in ours study is provided in the appendix of the paper. The two main independent variables GEN_DIV and HPPI represent genetic diversity and disease prevalence, respectively, and consistent with our formulated hypotheses we expect a negative and statistically significant coefficient estimates on the β parameter i.e. $\beta < 0$. As mentioned in the introduction, we use measures of expected heterozygosity provided by Ashraf and Galor (2013) to capture genetic diversity whereas we employ the 9-digit historical pathogen prevalence index found in Murray and Schaller (2010) to measure disease prevalence. Our estimated regressions also include a host of control variables used to address possible omitted variables bias. We present three main sets of controls commonly used in the deep roots literature. Firstly, we use geographic controls inclusive of land suitability which measures soil suitability for cultivation of agricultural crops (*land_suit*) collected from Michalopoulos (2012), average elevation (*elevation*), geographic latitude (*latitude*) and geographical longitude (*longitude*) collected from the CIA World factbook. Secondly, we use ecological controls inclusive of ecological fractionization (*eco_frac*) and ecological polarization (*eco_polar*) which are collected from (Fenkse, 2014) these indices measure the extent to which a population approximates a territory in which two vegetation types occupy half its area. Thirdly, we use average temperature (*temp*) and average precipitation (*precip*) as climate controls which we source from Harris et al. (2014).

As part of our sensitivity analysis, we employ an additional 4 sets of control and dummy variables. Firstly, we use state and antiquity (S&A) measures of intuitional quality which includes variables measuring i) the time of Neolithic transitions from hunting and gathering to agriculture based societies (*Neolithic*), as sourced from Putterman and Weil (2010) ii) The state history variable (*Stat_Hist*), also sourced from Putterman and Weil (2010) which measures the depth of experience with state institutions by capturing the strength of locally-dominated

government structures above tribal levels within a territorial geographic scope iii) the time elapsed since the original (uninterrupted) settlement of human populations provided by Ahlerup and Olsson (2012) (*Origtime*) iv) the 2019 per capita level of heath expenditure as measure of the quality of health institutions (*health*). Secondly, we employ legal origins dummies of La Porta et al. (1999) which categorize countries according to 5 legal systems (i.e. English, French, German, Scandinavian or Socialist). Thirdly, we use the natural resources dummy (oil and gas reserve dummy) which we source from Lujala et al. (2007). Lastly, we use the dummy variables for Islands as source form CIA World factbook. We consider this later dummy important since most countries which have recorded no cases of coronavirus are incidental Island economies. As previously mentioned, we also employ two-stage least squares (2SLS) estimators to address potential heterogeneity problem of endogeneity within the regressions. To this end we employ we employ two sets of instruments. For the first set of instruments, we follow Asharaf and Galor (2013) and use the migratory distance from Addis Ababa as an instrument for expected heterozygosity, which we then use to estimated COVID-19 infections and mortalities. For the second set of instruments, we follow Nikolaev and Salahodjaev (2017) and Ang (2019) and employ measures of collectivism/individualism from Hofstede (2001) as an instrumental variable for historical pathogen prevalence. The descriptive statistics of all the variables used in ours study is summarized in Table 1 in the appendix of the paper.

3. EMPIRICAL EVIDENCE

This section of the paper presents the OLS estimates of the regressions (1) and (2) which are summarized in Table 2. Panel A reports the results between genetic diversity and COVID_19 infections/mortalities (i.e. columns (1) - (8)) whilst Panel B reports the reports the results between disease prevalence and COVID_19 infections/mortalities (i.e. columns (9) - (16)). Columns (1), (5), (9) and (13) presents the models without any control variables which produce negative and statically significant estimates at all levels of significance. Generally, these findings support Lively (2010) hypothesis that populations with higher genetic diversity and more disease prevalence are associated with lower COVID_19 infections and mortalities. We are, however, concerned by the low R-squared values, particularly when genetic diversity is the independent variables (columns (9) – (16)) and the variable explains only between 2 and 6 percent of variation in COVID_19 morbidities and mortalities.

Suspecting omitted variables bias as the reason for these low explanatory power in the regressions, columns (2), (6), (10) and (14) present the models inclusive of four geographic controls (i.e. land suitability, average elevation, longitude and latitude) which significantly improves on all R-squared values which now explain between 28 and 48 percent of variation in the regressions. Moreover, we observe that geographic factors such as latitude and longitude are positively correlated with COVID infections and mortalities whilst land suitability is inversely correlated with the disease. These findings imply that countries further from the equator and Prime Meriden as well as those with less suitable cultivation land suffer from more COVID infections and mortalities.

In columns (3), (7), (11) and (15) we add two ecological controls (i.e. ecological fractionalization and polarization) whilst in columns (4), (8), (12) and (16) we add another two climate control variables (average temperature and precipitation). Notably the addition of the last two sets of controls does not offer much change in the magnitude of regression estimates for genetic diversity and disease prevalence and neither does it significantly improve the R-squared variable. We do, however, observe negative and statistically significant estimates on the temperate variable, which implies that areas with higher temperatures have lower COVID infections and deaths. Similar findings have been recently reported in O'Reilly et al. (2020) who observe a low survival rate of the SARS-CoV virus in geographical areas with higher temperatures and humidity levels. Nevertheless, we note that in all regressions the magnitude of the coefficient estimates on genetic diversity and disease prevalence variables is larger in absolute terms in comparison to coefficient estimates on other control variables, hence highlighting the dominance of evolutionary and behaviour factors in explaining movements in COVID-19 infections and mortalities.

Table 2: Baseline regressions

	Dependent variable: Log (SARS-CoV-2)								
Panel A	**Infections**					**Mortalities**			
Independent variable	**(1) No controls**	**(2) Add biogeography**	**(3) Add ecology**	**(4) Add Climate**		**(5) No controls**	**(6) Add biogeography**	**(7) Add ecology**	**(8) Add Climate**
Gen_div	-12.91 (4.18)***	-28.13 (4.17)***	-26.52 (4.20)***	-26.90 (4.52)***		-8.33 (4.15)***	-25.68 (4.73)***	-24.46 (4.67)***	-23.99 (5.03)***
Land_suitability		-0.02 (0.008)***	-0.12 (0.06)*	-0.13 (0.06)**			-0.02 (0.01)	-0.11 (0.07)	-0.11 (0.07)
Elevation		0.0004 (0.0005)	0.0002 (0.0006)	0.00003 (0.0006)			0.001 (0.0005)***	0.0009 (0.0005)*	0.0009 (0.0005)*
Latitude		0.07 (0.01)***	0.07 (0.01)***	0.05 (0.01)**			0.07 (0.01)***	0.07 (0.01)***	0.06 (0.01)***
Longitude		0.007 (0.004)*	0.007 (0.004)*	0.008 (0.004)**			0.007 (0.004)*	0.007 (0.004)*	0.007 (0.004)*
Dist_river		-0.0005 (0.0004)	-0.0005 (0.0004)	-0.0008 (0.0004)*			-0.0007 (0.0005)	-0.0008 (0.0005)	-0.0009 (0.0006)
Eco_frac			2.03 (1.25)	1.77 (1.25)				2.04 (1.46)	1.88 (1.53)
Eco_pol			-1.95 (1.31)	-1.43 (1.31)				-1.97 (1.29)	-1.81 (1.30)
Temp				-0.05 (0.03)*					-0.02 (0.03)
Precipitation_mean				-0.0005 (0.0003)					-0.0001 (0.0003)
Constant	15.97 (2.96)***	25.39 (2.88)***	24.54 (2.93)***	26.56 (2.99)***		9.68 (2.97)***	20.15 (3.28)***	19.56 (3.33)***	19.83 (3.28)***
R^2	0.06	0.48	0.49	0.52		0.03	0.40	0.41	0.42
Obs	143	127	127	127		143	111	111	111

Panel B									
independent variable	**(9) No controls**	**(10) Add biogeography**	**(11) Add ecology**	**(12) Add Climate**		**(13) No controls**	**(14) Add biogeography**	**(15) Add ecology**	**(16) Add Climate**
Disease_9	-1.75 (0.29)***	-1.19 (0.33)***	-1.24 (0.29)***	-1.24 (0.29)***		-1.29 (0.35)***	-0.81 (0.34)**	-0.82 (0.33)**	-0.83 (0.34)
Land_suitability		-0.05 (0.0)***	-0.22 (0.07)***	-0.17 (0.040***			-0.05 (0.01)***	-0.19 (0.08)**	-0.13 (0.05)***
Elevation		0.0003 (0.0006)	0.00005 (0.0007)	-0.0002 (0.0007)			0.001 (0.0005)**	0.0008 (0.0006)	0.0007 (0.0006)
Latitude		0.03 (0.01)***	0.04 (0.01)***	0.04 (0.009)***			0.03 (0.01)***	0.04 (0.01)***	0.04 (0.01)***
Longitude		-0.003 (0.004)	-0.002 (0.003)	-0.001 (0.003)			-0.003 (0.004)	-0.002 (0.003)	-0.001 (0.003)
Dist_river		-0.0007 (0.0005)	-0.0007 (0.0004)*	-0.001 (0.0005)*			-0.0009 (0.0005)*	-0.001 (0.0005)*	-0.001 (0.0006)*
Eco_frac			3.83 (1.47)**	5.41 (1.58)***				3.22 (1.73)*	5.12 (1.82)***
Eco_pol			-2.77 (1.40)*	-3.61 (1.56)**				-2.29 (1.51)	-3.67 (1.62)**
Temp				-0.01 (0.003)***					-0.01 (0.003)***
Precipitation_mean				-0.002 (0.003)					0.0001 (0.003)
Constant	6.99 (0.39)***	6.54 (0.45)***	6.46 (0.76)***	6.55 (0.80)***		3.81 (0.47)***	3.00 (0.48)***	2.86 (0.86)***	2.74 (0.86)***
R^2	0.20	0.34	0.39	0.42		0.12	0.28	0.30	0.34
Obs	143	127	127	127		124	111	111	111

Notes: "***", "**", "*" denote 10%, 5% and 1% critical levels, respectively. White heteroscedasticity-consistent standard errors reported in (). P-values are reported in [].

4. SENSITIVITY ANALYSIS: ADDITIONAL CONTROLS AND DUMMY VARIABLES

To ensure the robustness of our previous findings, this section of the paper presents a re-estimation of previous regressions after including more controls and dummy variables. Firstly, we include 3 sets of dummies for Islands, legal origins and natural resources and report these findings in columns (1), (5), (9) and (13) of Table 3. Secondly, in columns (2), (6), (10) and (14) we add controls for the number of years since the Neolithic transition from hunting and gathering to agriculture societies. Thirdly, in columns (3), (7), (11) and (15), we add controls for state history which accounts for the depth of experience with state institutions above tribal levels within a territorial geographic scope. Fourthly, in columns (4), (8), (12) and (16) we add controls for time elapsed since original (uninterrupted) human settlement and for health expenditure as a proxy for quality of current health institutions.

As can be collectively observed in Table 3, the inclusion of additional dummies and controls does not change the sign nor the magnitude of the genetic diversity and disease prevalence variables and yet we note an improvement in the explanatory power of all regressions (i.e. R^2). We also note that none of the state and antiquity (S&A) variables (i.e. Neolithic transition, state history and original time since human settlement) are significantly correlated with COVID_19 figures, implying that historical institutional advantages are non-detrimental towards COVID_19 infections and deaths. Moreover, the positive and significant coefficient estimate on current health institution variables reported in columns (4) and (16) further implies that economies with less advanced health institutions have been less affected by the COVID pandemic. This finding can be treated as additional evidence of factors other than technological and institutional factors being responsible for the differing patterns in distribution of COVID_19.

Table 3: Regressions inclusive of additional controls and dummies

	Dependent variable: Log (SARS-CoV-2)								
	Infections					**Mortalities**			
Independent variable	**(1) Add dummies**	**(2) Add Neolithic**	**(3) Add statehist**	**(4) Add Origtime+health**		**(5) Add dummies**	**(6) Add Neolithic**	**(7) Add statehist**	**(8) Add Origtime+health**
Gen_div	-26.97 (4.37)***	-27.39 (4.64)***	-27.36 (4.71)***	-14.57 (5.93)***		-25.21 (5.39)***	-25.38 (5.56)***	-25.35 (5.60)***	-16.41 (7.69)**
Neolithic		0.28 (0.37)	0.27 (0.37)	0.43 (0.34)			0.16 (0.39)	0.15 (0.41)	0.19 (0.41)
Statehist			0.01 (0.26)	0.03 (0.20)				0.03 (0.27)	0.10 (0.28)
Origtime				-0.20 (0.18)					-0.21 (0.29)
Health				4.11 (1.09)***					2.21 (1.34)
Ecology controls	Yes	Yes	Yes	Yes		Yes	Yes	Yes	Yes
Geography controls	Yes	Yes	Yes	Yes		Yes	Yes	Yes	Yes
Climate controls	Yes	Yes	Yes	Yes		Yes	Yes	Yes	Yes
Legal origin dummies	Yes	Yes	Yes	Yes		Yes	Yes	Yes	Yes
Island dummies	Yes	Yes	Yes	Yes		Yes	Yes	Yes	Yes
Natural resources dummy	Yes	Yes	Yes	Yes		Yes	Yes	Yes	Yes
Constant	22.53 (3.23)***	20.45 (4.70)***	20.50 (4.74)***	9.85 (5.54)*		16.59 (3.61)***	16.29 (4.38)***	16.41 (4.47)***	10.49 (5.58)*
R^2	0.62	0.61	0.61	0.66		0.57	0.58	0.58	0.60
Obs	127	117	117	115		111	105	105	103
independent variable	**(9) Add dummies**	**(10) Add Neolithic**	**(11) Add statehist**	**(12) Add Origtime**		**(13) Add dummies**	**(14) Add Neolithic**	**(15) Add statehist**	**(16) Add health**
Disease_9	-1.51 (0.39)***	-1.83 (0.44)***	-1.83 (0.44)	-1.27 (0.67)*		-1.25 (0.46)***	-1.54 (0.49)***	-1.56 (0.48)***	-1.08 (0.54)**

Neolithic		0.93 (0.76)	0.82 (0.67)	0.84 (0.60)			0.50 (0.39)	0.40 (0.40)	0.35 (0.37)
Statehist			0.29 (0.26)	0.27 (0.24)				0.27 (0.31)	0.27 (0.30)
Origtime				-0.51 (0.33)					
Health									3.23 (1.27)**
Ecology controls	Yes	Yes	Yes	Yes		Yes	Yes	Yes	Yes
Geography controls	Yes	Yes	Yes	Yes		Yes	Yes	Yes	Yes
Climate controls	Yes	Yes	Yes	Yes		Yes	Yes	Yes	Yes
Legal origin dummies	Yes	Yes	Yes	Yes		Yes	Yes	Yes	Yes
Island dummies	Yes	Yes	Yes	Yes		Yes	Yes	Yes	Yes
Natural resources dummy									
Constant	3.50 (1.26)***	3.49 (4.15)	2.13 (4.36)	2.59 (5.26)		1.09 (1.45)	3.83 (3.31)	2.62 (3.51)	4.12 (3.04)
R^2	0.58	0.58	0.58	0.60		0.52	0.55	0.55	0.58
Obs	127	117	117	117		111	105	105	103

Notes: "***", "**", "*" denote 10%, 5% and 1% critical levels, respectively. White heteroscedasticity-consistent standard errors reported in (). P-values are reported in [].

5. TWO-STAGE LEAST SQUARES (2SLS) ESTIMATES

So far, we have not addressed the issue of endogeneity in the estimated regressions. In this section of the paper, we use two-staged least squares (2SLS) model in which we use instruments variables for genetic diversity and disease prevalence. In applying the 2SLS estimators to examine the relationship between genetic diversity and COVID_19 infections/mortalities, we specify the first stage regression the one used in Ashraf and Galor (2013), Sequeira et al., (2019), Sequeira and Santos, (2019) and Arbatli et al., (2020) i.e.

$$GEN\_DIV_i = \mu + \beta\ MIGR\_DIST_i + CONTROLS_i + \varepsilon_i \quad (3)$$

And then extract the estimated values of genetic diversity from regression (3) and use them in the following second-stage regression:

$$SARS\text{-}CoV\text{-}2_i = \mu + \beta\ GEN\_DIV_i + CONTROLS_i + \varepsilon_i \quad (4)$$

Moreover, we follow Nikolaev and Salahodjaev (2017) and Ang (2019) who use disease prevalence as an instrument for collectivism/ individualism measures of psychological behaviour proposed by Gelfand et al. (2004). In applying these instruments in our study, we propose the following 2SLS estimation regressions. Under the first stage regression, we model individualism/collectivism as being endogenous to disease prevalence i.e.

$$IND\_COLL_i = \mu + \beta\ HPPI_i + CONTROLS_i + \varepsilon_i \quad (5)$$

And then we extract the estimates of in individualism/collectivism and model them as being exogenous towards COVID_19 infections and mortalities in the second-stage regression i.e.

$$SARS\text{-}CoV\text{-}2_i = \mu + \beta\ IND\_COLL_i + CONTROLS_i + \varepsilon_i \quad (6)$$

From regressions (3) - (6), we employ a set of controls for geography, climate, ecology, Islands, natural resources and institutions as used in the previous section of the paper. Table 4 presents a summary of the 2SLS results, with Panel A reporting the estimates of regressions

(3) and (4) whereas Panel B reports the estimates of regressions (5) and (6). Note that, as in the previous section of the paper, we present the estimates of our models in a stepwise fashion with columns (1), (5), (9) and (13) only including baseline controls and dummies (i.e. geography controls, climate controls, ecological controls, Islands dummies and natural resource dummies); columns (2), (6), 10) and (14) adding the number of years since the Neolithic transition; columns (3), (7), (11) and (15) adding State history; and lastly columns (4), (8), (12) and (15) adding time since original human settlement.

Based on the results reported in Panel A, we find that migratory distance is negatively and significantly related with genetic diversity in all first-stage regression estimators corresponding to columns (9) – (16) and we note that these findings are in alignment with those reported in Ashraf and Galor (2013), Sequeira et al., (2019), Sequeira and Santos, (2019) and Arbatli et al., (2020). From the second-stage estimates reported in columns (1) to (8), we observe familiar negative and statistical significant estimates on the on the 'instrumented' genetic diversity variable. These findings confirm a mechanism in which longer (shorter) migratory distance from the origin is negatively (positively) correlated with genetic diversity, which then becomes a positive (negative) predictor of COVID-19 infections and mortalities. One the other hand, the first stage estimates from Panel B reveal a negatively and statistically significant estimates between disease prevalence and collectivism as previously found in Nikolaev and Salahodjaev (2017) and Ang (2019), whereas the second stage estimates further reveal a negative correlation between the 'instrumented' collectivism variable and COVID-19 infections and mortalities. These findings confirm a mechanism in which populations with longer (shorter) historical experiences with pathogen diseases are associated with societies characterized by more (less) collectivism behaviour, which, in turn, is a negative (positive) predictor of COVID-19 infections and mortalities. This later mechanism reflects the mechanism described in the 5th Chapter of Diamond's (1997) book which suggests that populations with longer experiences with diseases tend to *"…[evolve] substantial resistance to new[er] disease…"* (Diamond 1997: 92).

Table 4: 2SLS estimates

Panel A:	**Dependent variable: Log (SARS-CoV-2)**								
2nd stage estimates	**Infections**					**Mortalities**			
	(1) Add dummies	**(2) Add Neolithic**	**(3) Add statehist**	**(4) Add Origtime+health**		**(5) Add dummies**	**(6) Add Neolithic**	**(7) Add statehist**	**(8) Add Origtime+health**
Gen_div	-26.97 (4.37)***	-27.39 (4.64)***	-27.36 (4.71)***	-14.57 (5.93)***		-25.21 (5.39)***	-25.38 (5.56)***	-25.35 (5.60)***	-16.41 (7.69)**
Neolithic		0.28 (0.37)	0.27 (0.37)	0.43 (0.34)			0.16 (0.39)	0.15 (0.41)	0.19 (0.41)
Statehist			0.01 (0.26)	0.03 (0.20)				0.03 (0.27)	0.10 (0.28)
Origtime				-0.20 (0.18)					-0.21 (0.29)
Health				4.11 (1.09)					2.21 (1.34)
Ecology controls	Yes	Yes	Yes	Yes		Yes	Yes	Yes	Yes
Geography controls	Yes	Yes	Yes	Yes		Yes	Yes	Yes	Yes
Island dummies	Yes	Yes	Yes	Yes		Yes	Yes	Yes	Yes
Natural resources dummy	Yes	Yes	Yes	Yes		Yes	Yes	Yes	Yes
Constant	22.53 (3.23)***	20.45 (4.70)***	20.50 (4.74)***	9.85 (5.54)*		16.59 (3.61)***	16.29 (4.38)***	16.41 (4.47)***	10.49 (5.58)*
R^2	0.62	0.61	0.61	0.66		0.57	0.58	0.58	0.60
Obs	127	117	117	115		111	105	105	103
IV F-statistic	10.66***	8.52***	7.99***	8.77***		7.23***	6.63***	6.21***	5.75***
1st stage estimates	**Dependent variable: Log (Gen_div)**								

Migra_dist	-0.007 (0.00004)***	-0.007 (0.00004)***	-0.007 (0.00004)***	-0.007 (0.00004)***		-0.007 (0.00004)***	-0.007 (0.00004)***	-0.007 (0.00004)***	-0.007 (0.00004)***
Controls and dummies	✔	✔	✔	✔		✔	✔	✔	✔
R^2	0.99	0.99	0.99	0.99		0.99	0.99	0.99	0.99
Panel B:									
2nd stage estimates									
	(9) Add dummies	**(10) Add Neolithic**	**(11) Add statehist**	**(12) Add Origtime+health**		**(13) Add dummies**	**(14) Add Neolithic**	**(15) Add statehist**	**(16) Add Origtime+health**
Coll_ind	-4.90 (2.23)**	-5.86 (2.63)**	-5.87 (2.73)**	-3.11 (1.75)*		-5.41 (2.66)**	-6.63 (2.93)**	-6.64 (3.04)**	-4.17 (2.15)*
Neolithic		2.95 (0.81)***	2.99 (1.30)**	2.43 (0.95)**			3.53 (0.97)***	3.60 (1.51)**	3.00 (1.32)**
Statehist			-0.05 (0.71)	0.41 (0.39)				-0.07 (0.80)	0.21 (0.55)
Origtime				-0.72 (0.39)*					-0.55 (0.48)
Health				2.73 (2.11)					2.18 (2.61)
Ecology controls	Yes	Yes	Yes	Yes		Yes	Yes	Yes	Yes
Geography controls	Yes	Yes	Yes	Yes		Yes	Yes	Yes	Yes
Island dummies	Yes	Yes	Yes	Yes		Yes	Yes	Yes	Yes
Natural resources dummy	No	No	No	Yes		No	No	No	Yes
Constant	9.16 (1.91)***	16.24 (6.64)***	16.65 (11.11)	7.96 (7.84)		5.54 (2.08)**	24.19 (7.94)***	24.78 (12.92)*	17.15 (10.98)
R^2	0.17	0.11	0.12	0.49		0.09	0.19	0.21	0.41
Obs	66	62	62	61		66	62	62	61

IV F-statistic	11.31***	16.66***	15.48***	20.73***		9.79***	16.76***	15.48***	20.93***
1st stage estimates	**Dependent variable: Log (Coll_ind)**								
Disease_9	0.68 (0.21)***	0.72 (0.24)***	0.70 (0.25)***	0.55 (0.26)***		0.68 (0.21)***	0.72 (0.24)***	0.70 (0.25)***	0.55 (0.26)***
Controls and dummies	✔	✔	✔	✔		✔	✔	✔	✔
R^2	0.54	0.57	0.57	0.60		0.54	0.57	0.57	0.60
Obs	66	62	62	61		66	62	62	61

Notes: "***", "**", "*" denote 10%, 5% and 1% critical levels, respectively. White heteroscedasticity-consistent standard errors reported in (). P-values are reported in [].

6. **CONCLUSIONS**

The COVID-19 epidemic has affected almost every sphere of human livelihood and hence interdisciplinary research is crucial towards understanding and providing possible solutions or guidelines in the world's battle with pandemic. Our study provides deeper knowledge on the susceptibility of different populations to the coronavirus disease by investigating whether natural selection evolutionary factors like genetic diversity and disease prevalence explain population vulnerability to the pandemic. Using statistical measures of expected heterozygosity and historical prevalence to infectious disease found in the 'deep roots literature', we find a negative and statistically significant co-relationship between these variables and total COVID-19 cases for a sample of 133 countries. We find that the magnitude of these variables as explanations for COVID-19 is superior to host of geographical, ecological, climate and institutional control factors. Instrumental variable estimates further reveal two transmission channels through which populations are susceptible to the COVID-19 virus. Firstly, there is the genetic route, where populations whose migratory distance from the origins of modern life are inversely related with genetic diversity, which then predicts movements in coronavirus infections and mortalities. Secondly, there is the behavioural channel, where populations with higher prevalence to infectious diseases develop 'collectivist' behavioural traits which acts as a antipathogen defence mechanism against the contraction and spread to the COVD-19 virus.

Our findings foremost hold relevant implications towards geneticists and related branches of medicinal research in their urgent need to develop a vaccine against the virus. Our observation of a negative co-movement between a population's expected heterozygosity and reported COVID-19 cases possibly reflects the variation in physiological immune system responses of different populations to the virus across the globe. Therefore, in developing a vaccine against the virus, geneticists may need to research on which specific human genes in populations are most resistant to the disease. Understanding how the variation in genetic selection of the COVID-19 pathogen towards different human populations would not only be useful towards formulating medication which can fight off the COVID-19 virus but also once such a vaccine is discovered and clinically-approved, then populations who are most vulnerable to the disease can be identified and prioritized for vaccination. Our results also have implications towards the behavioural governance of the coronavirus disease as a non-

pharmaceutical solution to fighting the pandemic. Our study particularly shows that societies which have previously gone through viral pandemics and have encouraged (discouraged) collectivism (individualism) behaviour are associated with lowest COVID-19 infections and mortalities. Therefore, governments worldwide need to incorporate behavioural approach in policy design which will not only control the spread of the infection but can present opportunities to speed-up the process of re-starting of economies. By replacing stringent lockdown strategies with behavioural adjustment strategies, economies can minimize the adverse economic repercussions of closing major sectors of the economy. This is more relevant for poorer countries who already face economic and social maladies such as high unemployment and extreme poverty and do not have access to necessary infrastructure and technology to protect themselves from the economic challenges presented by the virus. Without a clinically approved cure or treatment for the coronavirus disease, human survival may depend on our ability to adapt psychological and behavioural traits which will help avoid infections and manage contagion whilst simultaneously allowing people to live 'normal' lives.

REFERENCES

Acemoglu D., Johnson S. and Robinson J. (2001), "The colonial origins of comparative development: An empirical investigation", *American Economic Review*, 91(5), 1369-1401.

Adda J. (2016), "Economic activity and the spread of viral diseases: Evidence from high frequency data", *The Quarterly Journal of Economics*, 131(2), 891-941.

Alherup P. and Olsson O. (2012), "The roots if ethnic diversity", *Journal of Economic Growth*, 17, 71-102.

Anacleto O., Cabaleiro S., Villanueva B., Saura M., Houston R., Wooliams J. and Doeschl-Wilson A. (2019), "Genetic differences in host infectivity affect disease spread and survival in epidemics", *Scientific Reports*, 9, 12.

Ashar Q. and Galor O. (2012), "The "Out of Africa" hypothesis, human genetic diversity, and comparative economic development", *American Economic Review*, 103(1), 1-46.

Billing J. and Sherman P. (1998), "Antimicrobial functions of spices: Why some like it hot", *Quarterly Review of Biology*, 73, 3-49.

Cashdan E. and Steele M. (2013), "Pathogen prevalence, group bias, and collectivism in the standard cross-cultural sample", *Human Nature*, 24(1), 59-75.

Deshpande O., Batzoglou S., Feldman M. and Cavalli-Sforza L. (2008), "A serial founder effect model for human settlement out of Africa", *Proceedings of The Royal Society B*, 276, 291-300.

Fenske J. (2014), "Ecology, Trade, and states in pre-colonial Africa", *Journal of European Economic Association*", 12(3), 612-640.

Fincher C. and Thornhill R. (2008), "Assertive sociality, limited dispersal, infectious disease and the genesis of the global pattern of religious diversity", *Proceedings of the Royal Society B*, 275, 2587-2594.

Fincher C., Thornhill R. Murray D. and Schaller M. (2008), "Pathogen prevalence predicts human cross-culture variability in individualism/collectivism", *Proceedings of the Royal Society B*, 275, 1279-1285.

Fincher C. and Thornhill R. (2012), "Parasite stress promotes in-group assortative sociality: The cases of strong family ties and heightened religiosity", *Behavioral and Brain Sciences*, 35, 61-119.

Gallup J., Sachs J. and Mellinger A. (1999), "Geography and economic development", *International Regional Science Review*, 22, 179-232.

Gelfand M., Bhawuk D., Nishi L. and Bechtold D. (2004) "*Individualism and Collectivism*", In *Culture Leadership, and Organizations: the GLOBE study of 62 societies* (ed R. House, P. Hanges, M. Javidan, P. Dorfman and V. Gupta), pp. 437-512. Thousand Oaks, California, Sage Publications

Harris I., Jones P., Osborn T. and Lister D. (2014), "Updated high-resolution grids of monthly climatic observations – The CRU TS3.10 Dataset", *International Journal of Climatology*, 34, 623-642.

Jedwab R., Johnson N. and Koyama M. (2019), "Negative shocks an mass persecutions: Evidecne from the Black Death", *Journal of Economic Growth*, 24(4), 345-395.

Hofstede G. (2001), "*Culture's consequences. Comparing values, behaviors, institutions, and organizations across nations*", 2nd Editions, Thousand Oaks, California, Sage Publications.

Gelfand M., Bhawuk D., Nishii L. and Bechtold D. (2004), "*Individualism and collectivism*", In R. House, P. Hanges, M. Javidan, P. Dorfman and V. Gupta (Eds), Culture, Leadership and Organizations: The GLOBE study of 62 societies, Thousand Oaks, California, Sage Publications.

Kashima Y. and Kashima E. (1998), "Individualism, GNP, climate, and pronoun drop. Is individualism determined by affluence and climate, or does language use play a role?", *Journal of Cross Cultural-Psychology*, 34, 125-134.

Karlsson M., Nilsson T. and Pichler S. (2014), "The impact of the 1918 Spanish flu epidemic on economic performance in Sweden: An investigation into the consequences of an extraordinary mortality shock", *Journal of Health Economics*, 36, 1-19.

King K. and Lively C. (2012), "Does genetic diversity limit disease spread in natural host populations?", *Heredity*, 109(4), 199-203.

Koff W. and Williams M. (2020), "COVID-19 and immunity in aging populations – A new research agenda", *The New England Journal of Medicine*, (forthcoming).

La Porta R., Lopez-de-Silanes F., Shleifer A. and Vishny R. (1999), "The quality of government", *Journal of Law, Economics, and Organization*, 15(1), 222-279.

Lively C. (2010), "The effect of host genetic diversity on disease spread", *The American Naturalist*, 175(6), 149-152.

Lujala P., Rod J. and Thieme N. (2007), "Fighting over oil: Introducing a new dataset", *Conflict Management and Peace Science*, 24, 239-256.

Michalopoulos S. (2012), "The origins of Ethnolinguistic diversity", *American Economic Review*, 102, 1508-1539.

Miller A., Reandelar M., Fasciglione K., Roumenova V., Li Y. and Otazu G. (2020), "Correlation between universal BCG vaccination policy and reduced morbidity and mortality for COVID-19: An epidemiological study", Preprint at *medRxiv*, https://doi.org/10.1101/2020.03.24.20042937.

Murray D. and Schaler M. (2010), "Historical prevalence of infectious diseases within 230 geopolitical regions: A tool for investigating origins of culture", *Journal of Cross-Culture Psychology*, 41(9), 99-108.

Murray D., Trdeau R. and Schaller M. (2011), "On the origins of cultural differences in conformity: Four tests of the pathogen prevalence hypothesis", *Personality and Social Psychology Bulletin*, 37(3), 318-329.

Nikolaev B. and Salahodjaev R. (2017), "Historical prevalence of infectious diseases, cultural values, and the origins of economic institutions", *Kyklos*, 70(1), 97-128.

O'Reilly K., Auzenbergs M., Jafari Y., Liu Y., Flashe S. and Lowe R. (2020), "Effective transmission across the globe: The role of climate in COVID-19 mitigation strategies", *The Lancet Planetary Health*, 4(5), e172.

Phan T. (2020), "Genetic diversity and evolution of SARS-CoV-2', *Infection, Genetics and Evolution*", 81, 1-3.

Prugnolle F., Manica A. and Balloux F. (2005), "Geography predicts neutral genetic diversity of human populations", *Current Biology*, 15, 159-160.

Putterman L. and Weil D. (2010), "Post-1500 population flows and the long-run determinants of economic growth and inequality", *Quarterly Journal of Economics*, 125(4), 1627-2682.

Ramachandran S., Deshpande O., Roseman C., Rosenburg N., Feldman M., Cavalli-Sforza L. (2005), "Support from the relationship of genetic and geography distance in human populations for a serial founder effect originating from Africa", *Proceedings of the National Academy of Sciences of the United States*, 102, 15942-15947.

Sequeira T. and Santos M. (2019), "Technology in 1500 and genetic diversity", *Empirical Economics*, 56(4), 1145-1165.

Sequeira T., Santos M. and Ferreira-Lopez (2019), "human capital and genetic diversity", *Eurasian Economic Review,* 9(3), 311-330.

Suh E., Diener E., Oishi S. and Triandis H. (1998), "The shifting basis of life satisfaction judgments across cultures: emotion versus norms", *Journal of Personality and Social Psychology*, 74(2), 482-493.

Sibony A. (2020), "The UK COVID-19 response: A behavioural irony?", *European Journal of Risk Regulation*, 11(2), 1-8.

Toda A. (2020), "Early draconian social distancing maybe suboptimal in fighting the COVID-19 epidemic", VoxEu.org, 21st April 2020.

Uren C., Moller M., van Helden P., Henn B. and Hoal E. (2017), "Population structure and infectious disease risk in Southern Africa", *Molecular Genetics and Genomics*, 292(3), 499-509.

Wenham C., Smith J. and Morgan R. (2020), "COVID-19: the gendered impacts of the outbreak", *The Lancet*, 395(10227), 846-848.

Xiao Y. and Torok M. (2020), "Taking the right measures to control COVID-19", *The Lancet*, https://doi.org/10.1016/S1473-3099(20)30152-3.

Yang J., Zheng Y., Gou X., Pu K., Chen Z., Gou Q., Ji R., Wang H., Wang Y. and Zhou Y. (2020), "Prevalence of comorbidities and its effects in patients infected with SARS-CoV-2: A systematic review and meta-analysis", *International Journal of Infectious Disease*, 94, 91-95.

APPENDIX A

List of countries: Afghanistan, Albania, Algeria, Angola, Argentina, Armenia, Australia, Austria, Azerbaijan, Belgium, Benin, Bolivia, Botswana, Brazil, Brunei, Bulgaria, Burkina Faso, Burundi, Cambodia, Cameroon, Canada, Central African Republic (CAR), Chad, Chile, China, Colombia, Congo, Costa Rica, Cote d'Ivore, Croatia, Czech Republic, Denmark, Djibouti, Democratic Republic of Congo, Ecuador, Egypt, El Salvador, Equatorial Guinea, Eritrea, Estonia, Eswatini, Ethiopia, Finland, France, Gabon, Gambia, Georgia, Germany, Ghana, Greece, Guatemala, Guinea, Guinea-Bissau, Hungry, Iceland, India, Indonesia, Iran, Iraq, Ireland, Israel, Italy, Japan, Joran, Kenya, Kuwait, Laos, Latvia, Lebanon, Lesotho, Liberia, Libya, Lithuania, Luxembourg, Macedonia, Madagascar, Malawi, Malaysia, Mali, Mauritania, Mexico, Morocco, Mozambique, Namibia, Nepal, Netherlands, New Zealand, Niger, Nigeria, Norway, Oman, Panama, Peru, Philippines, Poland, Portugal, Romania, Russia, Rwanda, South Korea, Saudi Arabia, Senegal, Serbia, Sierra Leone, Slovakia, Slovenia, Somalia, South Africa, Spain, Sri Lanka, Sudan, Suriname, Sweden, Switzerland, Syrian, Tanzania, Thailand, Togo, Tunisia, Turkey, United Arab Emirates, Uganda, United Kingdom, Ukraine, Uruguay, United States of America, Uzbekistan, Venezuela, Vietnam, Yemen, Zambia, Zimbabwe.

APPENDIX B

Table 1: Summary statistics

Variable	Source	Obs	Mean	s.d.	Minimum	Maximum
Main dependent variables						
SARS-CoV-2 (infections)	John Hopkins Institute	134	7.15	2.43	1.61	13.34
SARS-CoV-2 (mortalities)	John Hopkins Institute	134	3.62	2.58	0.00	10.23
Main independent variables						

Gen_div	Ashraf and Golar (2013)	134	0.72	0.05	0.57	0.77
HPPI	Murray and Schaller (2010)	134	0.29	0.67	-1.78	1.20
Biogeography controls						
Land suitability	Michalopoulos (2012)	134	1.16	6.79	0.003	69.94
Average elevation	G-ECON project	128	562.18	475.11	0.64	2729.63
Latitude	CIA World Factbook	128	20.18	24.93	-42.00	64.00
Longitude	CIA World Factbook	128	19.90	51.98	-102.00	174.00
Distance to river	Gallup et al. (1999)	134	305.05	380.60	15.00	2385.58
Ecological controls						
Ecological fractionalization	Fenske (2014)	128	0.59	0.39	0.007	3.30
Ecological polarization	Fenske (2014)	128	0.67	0.19	0.02	0.91
Climate controls						
Temperature volume	Harris et al. (2014)	128	17.94	8.39	-4.94	28.55
Precipitation volume	Harris et al. (2014)	128	992.08	706.75	21.49	2933.82
Legal origins dummy	La Porta et al. (1999)	134	-	-	0	1
Island dummies	CIA's World Factbook	134	-	-	0	1
Natural resource dummy	Lujala et al. (2007)	129	-	-	0	1
Colony dummies	CIA's World Factbook	129	-	-	0	1
Historic institution controls						
Years since Neolithic revolution (Neolithic)	Putterman and Weil (2010)	118	10.50	16.39	5.99	145.60
State History (Statehist)	Putterman and Weil (2010)	118	-0.77	1.29	-3.57	7.50
Duration of human	Ahlerup and Olsson (2012)	134	10.44	1.98	-2.51	11.98

settlement (Origtime)						
2019 per capita expenditure on health	World Development Indicators	134	321356.82	605.25	22.89	6086.3
Instrumental variables						
Migratory distance from origin (Dist_orig)	Ashraf and Golar (2013)	118	7.29	1.82	0.00	11.81
Collectivism	Gelfand et al. (2004)	52	5.18	0.76	3.53	6.37