Possible Correlation between COVID-19 Contagion and Y-DNA Haplogroup R1b

Sebastiano Schillaci*

May 22, 2020 (updated on July 29, 2020)

Abstract

Here we develop some of the ideas we have first proposed in [1]. In particular, the linear correlation between the initial growth rate of COVID-19 contagion and the average Y-DNA haplogroup percentages in different countries is computed. In the case of haplogroup R1b, a positive correlation with high confidence level is found. Utilizing the maximum R1b percentages in place of the average ones, a more significant result is obtained. Considering an extended R1b data set, correlations with even higher confidence level are found (p-values 3.94×10^{-7} and 2.40×10^{-9}, respectively). Repeating the same procedure for the initial growth rate of deaths, similar results are obtained (p-values 9.17×10^{-11} and 2.18×10^{-12}, respectively). Furthermore, the correlation of haplogroup R1b with cases and deaths per capita is calculated over a five-month period, obtaining comparable results (e.g. p-value 2.45×10^{-17} on April 10th). The difference between the correlation with maximum R1b percentages and the correlation with average ones is decreasing over time. Finally, assuming the possible involvement of R1b carriers, three scenarios are outlined according to their passive or active role in the spread of the virus.

Contents

*EMAIL ADDRESS: sebastiano_schillaci@yahoo.com

1 Introduction

SARS-CoV-2 (from now on simply referred to as 'the virus') is the strain of coronavirus responsible for the ongoing pandemic of COVID-19, subsequently to a *spillover event*. Unless events of this kind happen, a pathogen that infects an animal is normally innocuous for animals belonging to a different species. Since different species are characterized by different gene pools, genetics is obviously involved in the susceptibility to a communicable disease. Even individuals from the same species can have a different susceptibility to a pathogen because of differences in their genotype. For example, individuals with a genetic mutation known as *CCR5-Δ32* show HIV resistance [2].

Also for COVID-19 transmission, it seems reasonable to suppose that the role of genetics is not negligible. Consider for example the apparently very different impact of COVID-19 on the Republic of Haiti and the Dominican Republic, sharing the island of Hispaniola and comparable on many respects (climate, population size, etc.), but otherwise populated by very different ethnic groups. Notwithstanding insufficient containment measures, poorer hygienic condi-

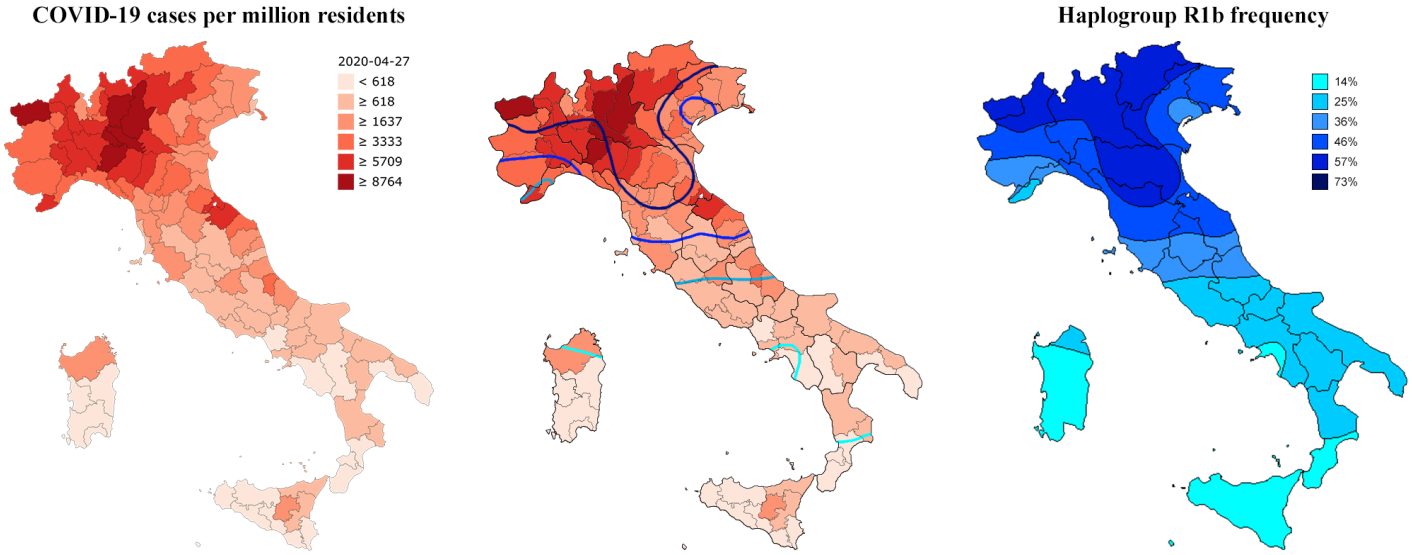

Figure 1: Confirmed COVID-19 cases per million residents in Italy on April 27th [8] (*left*), haplogroup R1b frequency in Italy [9] (*right*) and overlay of the two maps for an easier comparison (*center*).

tions and higher population density, Haiti was much less severely affected than the Dominican Republic so far [3]. Thus, it seems to be a sensible idea to look for possible genes related to COVID-19 transmission. Indeed such studies are already underway, but it can take a relatively long time to get the first results (see e.g. [4]).

It has been noted in past studies ([5], still on HIV) that, sometimes, progression and outcome of a virus infection are different among individuals belonging to different haplogroups. A haplogroup is a combination of alleles at different chromosomal regions that are closely linked and that tend to be inherited together. In human genetics, the most commonly studied haplogroups are Y-chromosome (Y-DNA) haplogroups, passed along the patrilineal line, and mitochondrial DNA (mtDNA) haplogroups, passed along the matrilineal line. They change only by chance mutation at each generation, with no intermixture between parents' genetic material. Only Y-DNA haplogroups will be considered from now on.

Haplogroups can be named either by the 'lineages' or by the 'mutations' that define the lineages themselves. In the first case, a capital letter identifying the major clade is followed alternatively by numerals and lower-case letters that hierarchically identify all subsequent subclades. In the second case, the same capital letter is followed by a dash and the name of the terminal mutation that defines the haplogroup [6]. For example, in modern populations, haplogroup R1 (alias R-M173) appears to be comprised of subclades R1a and R1b (also known as R-M420 and R-M343, respectively).

When the virus spread in Northern Italy and the first cities went under lockdown, many Southern people working and studying there, started going back to their homes. At that point it was expected the virus would have spread quickly also in the Southern regions, which did not happen. In Africa too, a sharp rise of COVID-19 cases was expected but this prediction did not come true either. This could be due to underlying genetic reasons. Comparing the haplogroup maps, it is apparent that in Southern Italy and Africa there is a low R1b frequency. Furthermore, the COVID-19 distribution map seems to correspond to the haplogroup R1b map. This work aims to give a quantitative measurement of this visual correlation.

2 Geographical distribution

Italy was the first country in Europe to be sensibly affected by the virus. For historical reasons, the genetic profile of the Italian population is geographically stratified [7]. In particular, the distribution of the haplogroup R1b varies greatly between Northern and Southern Italy, in a similar way to the distribution of COVID-19 cases (*Figure 1*).

While in Italy this correlation seems to be very clear, the situation tends to be more blurry in most of the other countries. Besides, it is very difficult to reliably compare the virus impact on different countries, mainly because of different testing and reporting policies (*Figure 2*). In any case it is quite evident that Southern Italy, Greece, Portugal and Africa were much less affected by the virus than it was expected, even taking into account the early lockdowns. These areas have a relatively low density of haplogroup R1b in the population, compared to most European countries (*Figure 3*). In Asia, a high concentration of haplogroup R1b is found in various ethnic groups: the Bashkirs in Bashkortostan (close to the Ural Mountains), the Lurs in Iran and the Turkmens in Turkmenistan (close to the Caspian Sea). The last one unfortunately does not disclose COVID-related data.

For a schematic description of the 'evidence' of the correlation in different countries and a more complete visual comparison 'side by side' refer to [1]. For a more detailed but complementary account see also [12] and [13].

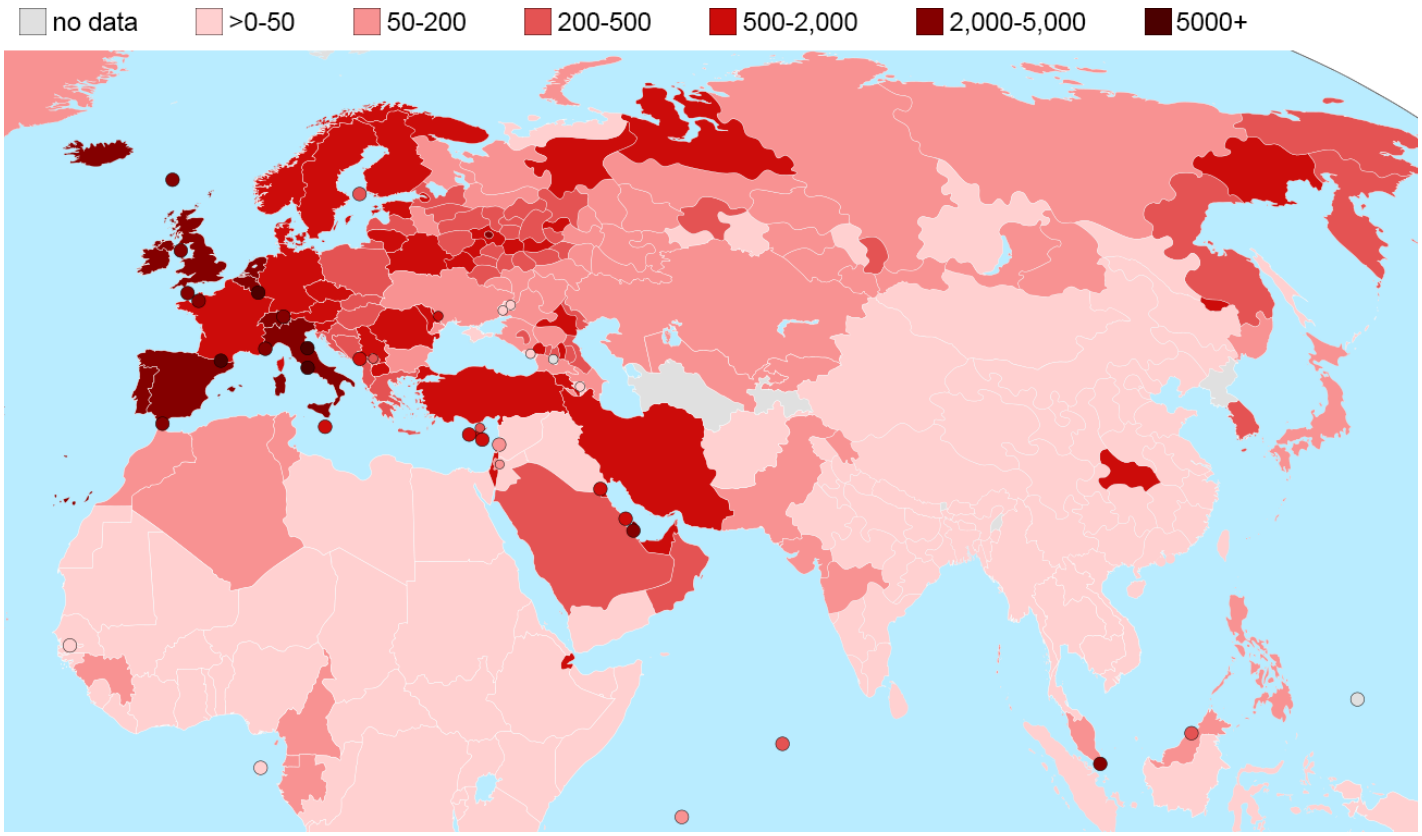

Figure 2: Confirmed COVID-19 cases per million inhabitants in Eurasia and part of Africa on April 27th [10].

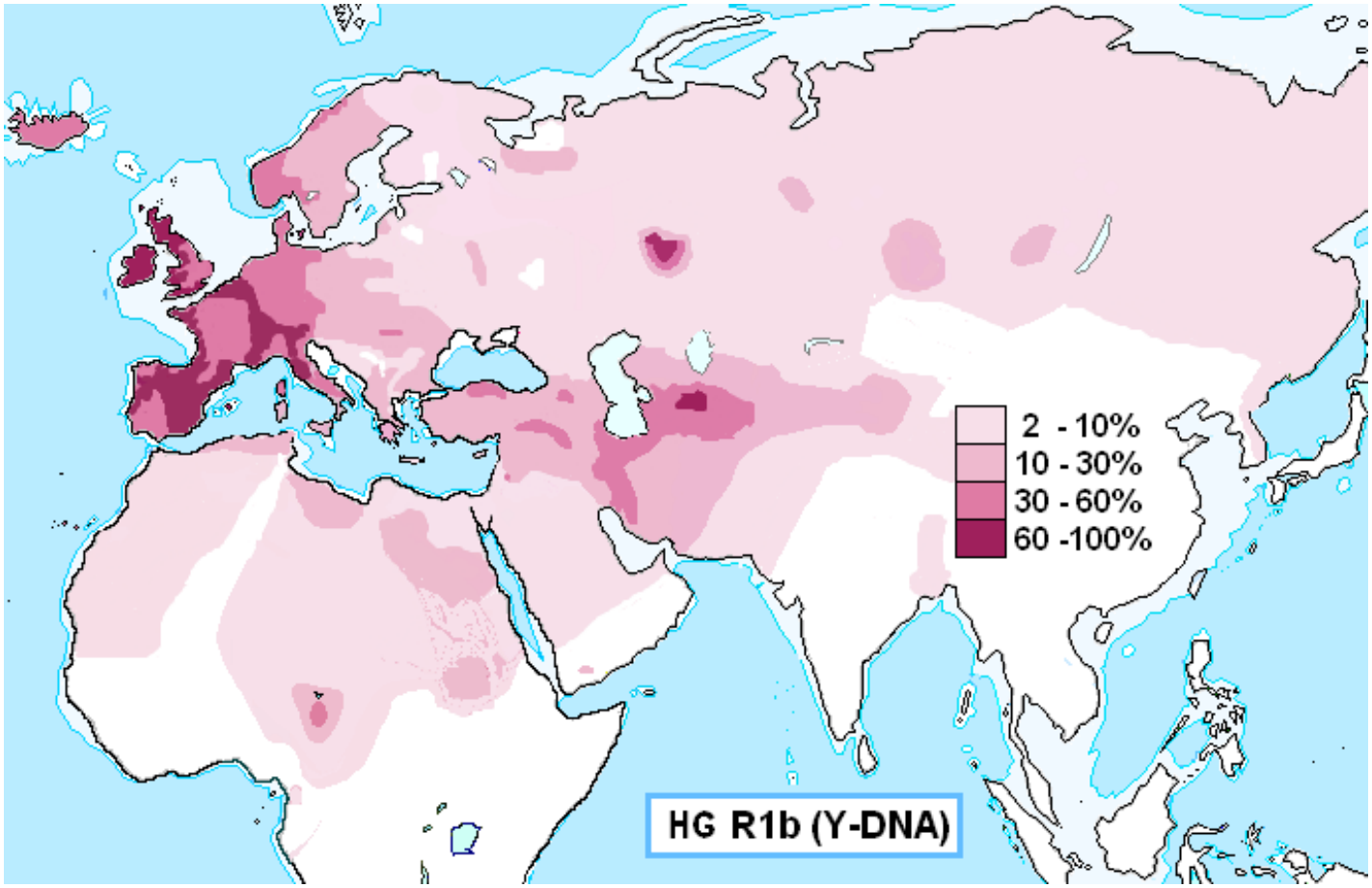

Figure 3: Haplogroup R1b distribution in Eurasia and part of Africa [11].

HAPLOGROUP	I	R1a	R1b	G	J2	J1	E	T	L	Q	N	Max R1b
a) Cases												
RMSE	0.075985	0.076046	**0.064448**	0.075954	0.074861	0.070816	0.07531	0.074249	0.076045	0.072944	0.075877	**0.059862**
CORR. COEFF.	0.055137	-0.037967	**0.53177**	0.061933	-0.17973	-0.36614	-0.14373	-0.21922	0.038038	0.28499	-0.076649	**0.61744**
p-VALUE	0.6865	0.78116	$\mathbf{2.4702 \times 10^{-5}}$	0.65023	0.18501	0.0055159	0.29059	0.10452	0.78076	0.033258	0.57447	$\mathbf{4.0044 \times 10^{-7}}$
b) Deaths												
RMSE	0.100236	0.101042	**0.078894**	0.101017	0.095832	0.094985	0.101244	0.099022	0.09995	0.101741	0.10108	**0.07737**
CORR. COEFF.	0.17136	-0.11703	**0.63142**	-0.11904	-0.33583	-0.35833	-0.098738	-0.22964	-0.18679	-0.0013959	-0.11379	**0.64938**
p-VALUE	0.20666	0.39034	$\mathbf{1.811 \times 10^{-7}}$	0.3822	0.011393	0.0066938	0.46907	0.08865	0.16805	0.99185	0.40371	$\mathbf{6.1572 \times 10^{-8}}$

Table 1: Linear regression of haplogroup percentages with growth rate of cases and deaths (sample $\mathcal{A}$).

3 Data sets

Data for the countries are taken from different sources: COVID-19 total [14] and sex-disaggregated [15] cases and deaths, distribution of European Y-DNA haplogroups [16] and haplogroup R1b [16] [17] in percentage. To minimize the risk of selection bias, no country is discarded if sufficient data are available. Combining the previous data sets, four samples are obtained:

- sample $\mathcal{A}$: 56 countries (total cases and deaths with European haplogroups);
- sample $\mathcal{B}$: 84 countries (total cases and deaths with haplogroup R1b);
- sample $\mathcal{C}$: 44 countries (male cases with European haplogroups);
- sample $\mathcal{D}$: 64 countries (male cases with haplogroup R1b).

Correlations could be assessed also for the 20 Italian regions, but the sample size would be too small to obtain significant results. The software MATLAB®, developed by MathWorks®, is used for the analysis.

4 Correlation with growth rate of cases and deaths

Obviously lockdown policies are crucial to determine the spreading rate of the virus, so we try to consider the initial speed of contagion, presumably before any significant containment measures were enforced. To this end, we apply the approach of [18] to sample $\mathcal{A}$.

In order to capture the initial exponential growth, we consider 12-day-long data sets starting from the first day in which a given country reached 30 cases (or the closest to 30, the last day if repeated). Those data are then fit for each country with a simple exponential curve $N_0 e^{\alpha t}$, where t is the number of days; the function `fit(x,y,'exp1')`, which minimizes the residuals by non-linear least squares regression, is used. Next, we perform a simple linear regression (with an estimated intercept term) to assess the relationship between the growth rate of contagion α and the haplogroup percentages. The p-values and the sample correlation coefficients, that estimate the population Pearson's linear correlation coefficients, are computed. In this case, the coefficient of determination R^2 is simply the square of the sample correlation coefficient. Conventionally, a p-value ≤ 0.05 (or sometimes even ≤ 0.01) is assumed as statistically significant. In the case of haplogroup R1b a positive correlation with highly significant p-value and $R^2 = 0.28278$ results (*Table 1a*).

The speed of contagion could vary greatly if the susceptible individuals are not spread evenly in the population. For example, if the basic reproduction number $R_0 = 2.5$, 60% of the population of a country getting immunized is the critical threshold over which the epidemic will peter out, because $2.5 \cdot (1 - 0.6) = 1$ [19]. But if the immunized individuals are segregated in one part of the country, the virus will still spread unencumbered in the remaining 40% of the population. This could be the case of Italy that, even having a much lower average percentage of haplogroup R1b than Spain, has faced an initial contagion almost as fast. So, probably, to calculate the correlation, instead of using the average percentage of haplogroup, it could make more sense to use the maximum one. For example, if we substitute the average Italian R1b percentage (39%) with the one in Lombardy (59%) [20], the linear correlation coefficient increases while the p-value decreases. In fact, substituting all the average R1b percentages with the maximum ones among their regions (see [16], [20], [21] and [22]), we obtain a higher correlation coefficient with $R^2 = 0.38123$ and a much smaller p-value (*Table 1a, last column*).

In the worst outbreaks many cases can go untested, especially if asymptomatic, so the number of deaths is considered a more reliable indicator of contagion than the number of cases. On this account, the previous method could also be applied to the number of deaths, obviously choosing a smaller number as a reference point. For example, to calculate the growth rate of deaths α, we can consider a 12-day-long data set starting from the first day in which a given country reached 2 deaths. Performing a simple linear regression to assess the relationship between these new rates α and the haplogroup percentages, even more significant results are obtained (*Table 1b*).

To increase the confidence level, the correlation of the growth rate of cases and deaths with R1b percentages is assessed also for the extended sample $\mathcal{B}$. The results exceed

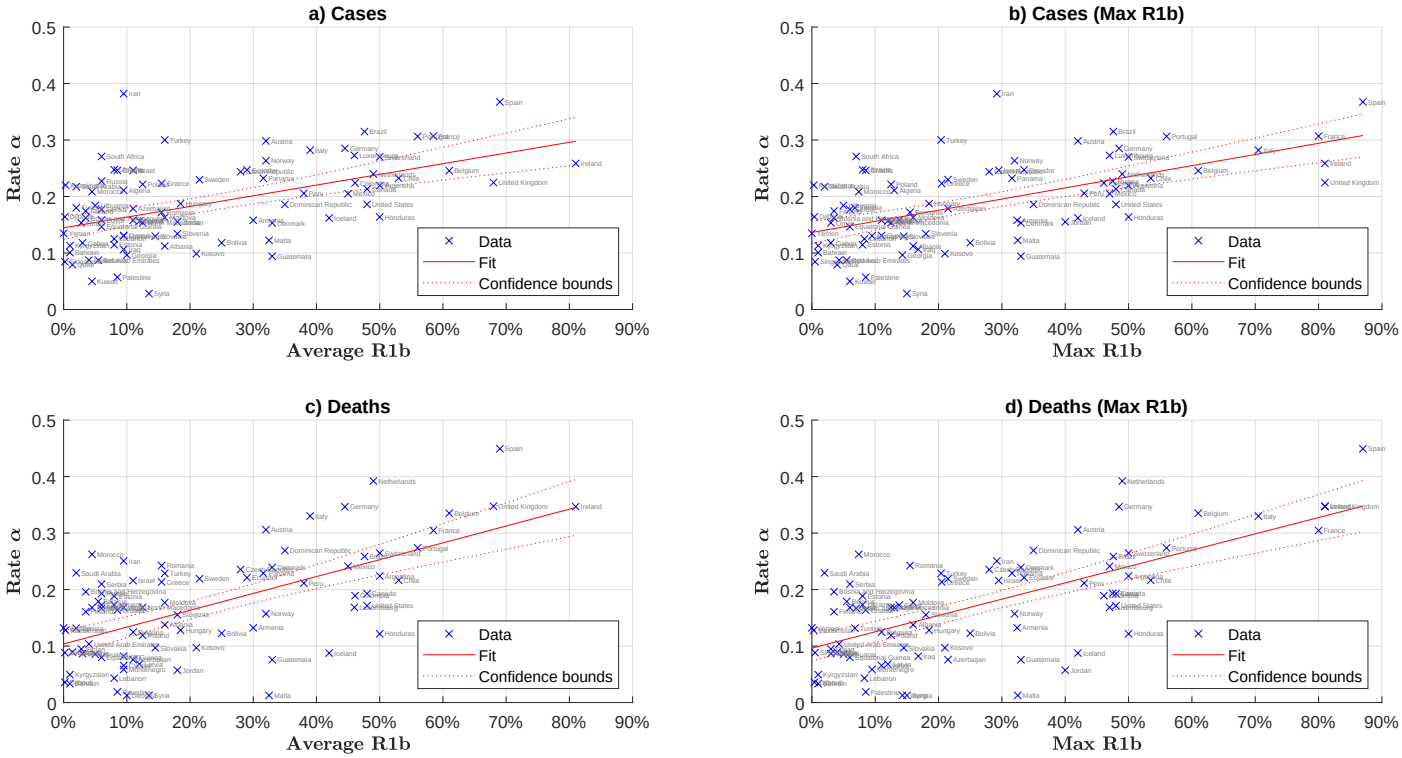

Figure 4: Growth rate α of cases and deaths vs. R1b percentage (sample $\mathcal{B}$) with 95% confidence interval.

	Cases		Deaths	
HAPLOGROUP	R1b	Max R1b	R1b	Max R1b
RMSE	0.062361	0.058689	0.072983	0.069777
CORR. COEFF.	0.52022	0.59498	0.63446	0.67371
p-VALUE	3.9389×10^{-7}	2.398×10^{-9}	9.1713×10^{-11}	2.1774×10^{-12}

Table 2: Linear regression of haplogroup percentages with growth rate of cases and deaths (sample $\mathcal{B}$).

5σ significance (*Table 2*).

It is apparent that all countries with high R1b percentage have had a high growth rate of the contagion (*Figure 4*). In countries with low R1b percentage the influence of the other haplogroups is not negligible, therefore the effects are less clear-cut. Besides, data for countries with low α or low R1b percentage may be proportionally more noisy. Data for low GDP countries could also correlate less well because of lack of testing or late testing. However the situation of the outliers should be examined on a case-by-case basis.

In fact, this correlation works particularly well in Western Europe, where haplogroup R1b is the most frequently occurring paternal lineage. One possible reason could be that a viral strain which is more contagious for individuals belonging to this haplogroup has got selected. The strain carrying the mutation *D614G*, which began spreading in Europe in early February, would probably be a good candidate [23].

Another problem is that the method used to calculate the rates α may be slightly biased. By the time countries with small populations reach 30 cases, they may have already enforced containment measures, which could result in underestimated rates α. This can also be verified performing a simple linear regression with rates α and population size, which gives significant results (on sample $\mathcal{B}$ $R = 0.33561$, p-value $= 0.001803$ for cases and $R = 0.27795$, p-value $= 0.01047$ for deaths). To eliminate this confounding factor, one solution could be to choose the starting day of the data set using, as a reference point, a number of cases (or deaths) proportional to the population size. Another possibility is to use the maximum value of α computed varying the starting day over the entire infection period. These solutions could be developed in future works.

5 Correlation with cases and deaths per capita

A somehow complementary approach could be to consider a later period, when all the countries have already enforced containment measures. At this time, it seems more appropriate to just compare the haplogroup percentages with the number of cases per capita for each country, like we did visually at the beginning. The correlation is assessed on sample $\mathcal{A}$ in order to have a set of countries with comparable starting times of the epidemic. For example, using 2020 world population estimates [24] [25], we can assess the

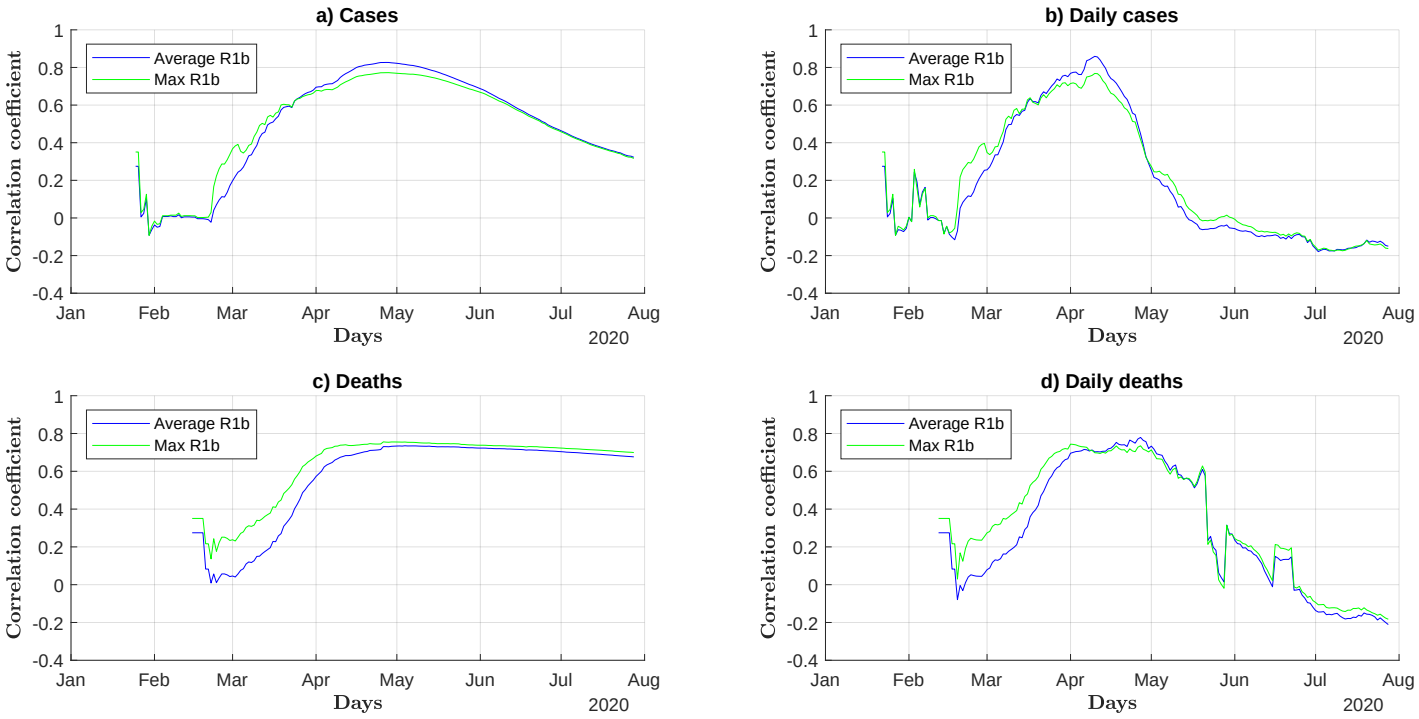

Figure 5: Correlation of haplogroup R1b percentages with COVID-19 cases and deaths per capita over time (sample $\mathcal{A}$).

Haplogroup	I	R1a	R1b	G	J2	J1	E	T	L	Q	N	Max R1b
a) Cases												
RMSE	0.00122	0.001227	**0.000698**	0.001238	0.001186	0.00116	0.001165	0.001219	0.001222	0.001239	0.001236	**0.000788**
Corr. coeff.	0.1794	-0.14624	**0.82671**	-0.053858	-0.29169	-0.35422	-0.34249	-0.18542	-0.16787	0.041448	-0.080526	**0.77229**
p-value	0.18583	0.28215	$\mathbf{4.2148 \times 10^{-15}}$	0.69341	0.029163	0.0073982	0.0097717	0.17126	0.21621	0.76166	0.55521	$\mathbf{3.1651 \times 10^{-12}}$
b) Deaths												
RMSE	0.000134	0.00013	**0.000092**	0.000134	0.000132	0.00013	0.000131	0.000133	0.000132	0.000134	0.000133	**0.000088**
Corr. coeff.	0.044933	-0.2547	**0.73048**	0.014727	-0.17132	-0.24434	-0.20184	-0.098695	-0.14912	-0.061148	-0.11494	**0.75431**
p-value	0.74228	0.05817	$\mathbf{1.6595 \times 10^{-10}}$	0.91421	0.20677	0.069544	0.13576	0.46926	0.27268	0.65438	0.39893	$\mathbf{1.9126 \times 10^{-11}}$

Table 3: Linear regression of haplogroup percentages with cases and deaths per capita on April 27th (sample $\mathcal{A}$).

correlation with the number of cases on the day *Figures 1* and *2* refer to (*Table 3a*).

We find again a highly significant correlation with haplogroup R1b and no significant association with the other ones. It is interesting to study how the value of the correlation with average and maximum R1b percentages varies over time (*Figure 5a*).

In the first weeks, the values are not significant because of the limited diffusion of the virus across the countries in the sample. After reaching a peak, they decrease, probably as an effect of the different containment policies adopted among countries. Moreover, the correlation with the maximum percentages is stronger than the correlation with the average ones at the beginning, and weaker afterwards. Such behavior could be explained with the hypothesis that the effects of contagion become evident first in areas with higher R1b frequency. Once the virus has spread more evenly across the countries, the effects of the average R1b frequency are more relevant.

To avoid the delay due to the cumulative count of cases, we can study the correlation with the daily number of cases per capita, averaged on a week to reduce the weekly fluctuations (*Figure 5b*). First the maximum percentages prevail over the average ones, then the situation get reversed and lastly the maximum percentages prevail again. Lockdowns, severing the connections between people, create a situation arguably more similar to an even distribution of R1b carriers. As a consequence, this behavior could also be related to the onset of lockdowns and to their lifting. Notably, a very strong correlation with average percentages results on April 10th ($R = 0.85909$, p-value $= 2.4514 \times 10^{-17}$).

In the last observed days the correlation with R1b becomes negative. The reason could be that the countries with higher R1b percentages were struck earlier than the others. Now they are in the descending phase of the contagion. On the contrary, in countries with lower R1b percentages where the virus has begun spreading later, the number of new cases is still growing.

As before, we can apply the same approach to the number of deaths per capita (see also [17]) and calculate the corre-

Experimental data	Scenario		
	1st	2nd	3rd
Positive correlation between R1b and cases	✓	✓	✓
Correlation of R1b with deaths stronger than with cases	✓	✓	✓
Negative correlation between R1b and male cases	–	–	✓
Negative correlation between R1b and cases, after the lockdown	–	–	✓
Minorities dying disproportionately	✗	✗	✓
Higher male lethality	✓	✓	–

Table 4: Compatibility matrix between experimental data and scenarios.

lation with haplogroup percentages, still on the same day (*Table 3b*).

We notice once more a significant correlation only with haplogroup R1b. In this case too, if we plot the correlation with average and maximum percentages of haplogroup R1b over time (*Figure 5c*), the difference between them tends to progressively decrease. Nevertheless, the correlation with the maximum percentages is always stronger than with the average ones. This is probably due to the delay of a couple of weeks between case detection and decease and to the decreasing case fatality rate [26], which makes the total death count change proportionally slower than the total case count. The latter statement is confirmed by the study of the correlation with the daily number of deaths per capita, still averaged on a week (*Figure 5d*). It is apparent from the graph that in the last observed period the correlation is not significant anymore (the two fluctuations are due to Spain death recounts on May 25th [27] and June 19th [28]).

6 Possible interpretations

Of course correlation does not imply causation and indeed there are many environmental factors that can help the virus to spread and there could exist environmental factors that favored the haplogroup R1b diffusion (e.g. like malaria selected the 'favism gene' [29]). But it seems unlikely there could be one environmental factor that does both, because of the very different timescales it should operate on (R1b emergence dates back to at least 14,000 years ago [30]).

Assuming the existence of an underlying genetic factor, one possibility would be that the interested genes are located on the sex chromosomes, considering the higher lethality for men [31]. They could be directly on the Y chromosome, being R1b a Y-DNA haplogroup, or even on the X chromosome ('unguarded', like in the case of color blindness [32]), especially since the ACE2 gene is located on the X chromosome and the human ACE2 receptor is known to be involved in the transmission of the virus [33].

Indeed, the difference in lethality between sexes is correlated with the haplogroup R1b frequency in the population, but this probably happens because they are both correlated with the case fatality rate [15]. One alternative explanation for the Y chromosome signal is that we are tracing an autosomal locus associated with the haplogroup R1b, like the ones identified in recent studies [34] [35] [36]. Unless differently specified, we will loosely refer also to those people as 'R1b carriers'.

Three scenarios, not necessarily mutually exclusive, seem to be compatible with our findings (*Table 4*). An R1b carrier could even experience all of them, at different stages of the disease.

In the first scenario, infected R1b carriers tend to develop very severe symptoms and therefore are more likely to be detected, resulting in a higher number of cases where the haplogroup frequency is higher. In this case, people belonging to the other haplogroups could be infected as well, with few or no symptoms, making them harder to discover.

In the second scenario, R1b carriers are more susceptible to infection. It could be even hypothesized that the more an individual is susceptible to infection, the higher the risk of dying if infected, which seems to hold true at least for children and old people [37] [38]. This could be another reason why the correlation of R1b with deaths is stronger than with cases. It could also help to explain why in Italy, initially, a very high case fatality rate was faced [39], but lately the virus seems to have lost strength [40], considering that the Italian population is genetically stratified. After the lockdown, with the virus less widespread and having been the more susceptible individuals already infected, it is less likely for the virus to encounter new clusters of susceptible individuals.

Probably in a similar way, the virus was already spreading earlier in Italy [41], since December according to [42], undetected. In this initial stage, the virus could have gone unnoticed because it spread among less susceptible individuals at a slow pace, because of the low (effective) reproduction number, and probably killing few people. Only when it encountered a susceptible community, the infection quickly grew and got noticed. Since the initial spreading rate correlates better with maximum R1b percentages than with average ones, it can be supposed that similar dynamics were at work also in other countries.

In the third scenario, R1b carriers are more contagious, being less symptomatic, having higher viral shedding, or both. Indeed, asymptomatic carriers [43] [44] and superspreaders [45] seem to play an important role in COVID-19 pandemic. A higher viral shedding may produce in the exposed individuals a higher viral load and, therefore, more severe infections [46]. This could be another reason why the correlation of R1b with deaths is stronger than with cases.

In areas where there is a higher R1b frequency, they could

spread the contagion, affecting people belonging to other haplogroups. This could be one of the reasons why black people and other ethnic minorities are dying in disproportionate amounts in UK [47] and US [48], where they live intermixed with R1b carriers, but not in Africa or Asia. It could also be one of the reasons why the correlation with R1b percentage becomes negative in the last period of observation. Asymptomatic carriers are less likely to be counted as cases and, once social distancing measures are enforced, they do not spread the virus either.

Moreover, because of insufficient data, we cannot exclude a prevalence of males among asymptomatic carriers. Thus, the involved genes could even be located on a sex chromosome. In particular, the Y chromosome can regulate the expression of various genes of the autosomes and X chromosomes [49], and males belonging to haplogroup R1b possibly have more effective immune and inflammatory responses [50]. This could make infected R1b carriers more likely to be asymptomatic, or paucisymptomatic. Besides, there is a significant negative correlation between the percentage of males over the total cases and the R1b percentage ($R = -0.40884$, p-value $= 0.0058619$ on sample $\mathcal{C}$; $R = -0.38441$, p-value $= 0.0017122$ on sample $\mathcal{D}$). Considering the positive correlation with the number of cases per capita, this appears to imply that R1b carriers are asymptomatic and, therefore, more difficult to detect.

In order to validate one or more of the aforementioned scenarios, a study of the genetic profile of the population in the most affected areas would be needed, such as the study in Vo' (Veneto) [4].

7 Conclusion

In this work, first, a coefficient α representing the growth rate of the contagion is obtained for 56 countries, using an exponential fit of the number of cases in the first days of spreading. Then, the Pearson's linear correlation coefficient between the rates α and the average Y-DNA haplogroup percentages in the corresponding countries is calculated, obtaining a highly significant correlation with haplogroup R1b. Lastly, the correlation between the same rates α and the maximum haplogroup R1b percentages in the corresponding countries is calculated, obtaining an even more highly significant correlation. A similar calculation is repeated with the number of deaths in the initial days of spreading, obtaining again highly significant correlation with haplogroup R1b. The same procedure is then repeated with an extended R1b data set of 84 countries, strengthening the results.

Moreover, it has been studied how the correlation of average and maximum haplogroup R1b percentages with cases per capita varies over a five-month period. The same procedure has been applied to the deaths per capita. Comparing the results, it emerges that at the beginning the correlation is stronger with the maximum R1b percentages than with the average ones, but over time the difference decreases. This could be explained supposing that the effects of contagion become evident first in areas with higher R1b frequency.

Three possible scenarios are then outlined, according to the passive or active involvement of R1b carriers in the diffusion of the virus. They could develop more severe symptoms, be more susceptible to infection, be more contagious, or one of their combinations. In particular, R1b carriers being asymptomatic (or pre-symptomatic) spreaders seems to be the most likely option.

Given those results, it could be useful to take into consideration the ideas here presented when confronting the current pandemic struggle. Our findings should warrant further epidemiological observational studies, such as case-control studies, to confirm or disprove the possible involvement of R1b carriers in the spread of the virus. A positive result could speed up the discovery of a treatment, help to make more reliable quantitative forecasting models or, at least, help to better tune the social distancing measures and to inform future vaccination campaign priorities.

The correlation data set is included as supplementary material (samples $\mathcal{A}$ and $\mathcal{B}$).

References

[1] Schillaci, S. *Possible link between COVID-19 susceptibility and genetic factors*. SXS. **April 13, 2020**. http://sxs.altervista.org/coronavirus/

[2] Samson, M., Libert, F., Doranz, B. J., et al. *Resistance to HIV-1 infection in Caucasian individuals bearing mutant alleles of the CCR-5 chemokine receptor gene*. Nature, 382(6593), 722–725. **August 1996**. https://doi.org/10.1038/382722a0

[3] Sieff, K. *Haiti, spared a major coronavirus outbreak so far, now a 'tinderbox' set to 'explode'*. The Washington Post. **May 15, 2020**. https://www.washingtonpost.com/world/the_americas/coronavirus-haiti-dominican-republic-hispaniola/2020/05/14/a51d0664-947f-11ea-87a3-22d324235636_story.html

[4] Suman, F. *Covid-19, ecco cosa spiega lo studio effettuato a Vo' (in Italian)*. Il Bo Live UniPD. **April 22, 2020**. http://ilbolive.unipd.it/it/news/studio-crisanti-coronavirus-vo

[5] Sezgin, E., Lind, J. M., Shrestha, S., et al. *Association of Y chromosome haplogroup I with HIV progression, and HAART outcome*. Human Genetics, 125(3), 281–294. **April 2009**. https://doi.org/10.1007/s00439-008-0620-7

[6] The Y Chromosome Consortium. *A Nomenclature System for the Tree of Human Y-Chromosomal Binary Haplogroups*. Genome Research, 12(2), 339–348. **February 2002**. https://doi.org/10.1101/gr.217602

[7] Hay, M. *Genetic history of the Italians*. Eupedia. **December 2017**. https://www.eupedia.com/genetics/italian_dna.shtml

[8] Wikimedia Commons contributors. *COVID-19 outbreak Italy per capita cases map*. Wikimedia Commons. **April 28, 2020**. https://upload.wikimedia.org/wikipedia/commons/archive/e/e6/20200501073349!COVID-19_outbreak_Italy_per_capita_cases_map.svg

[9] Wikimedia Commons contributors. *Frequenza dell'aplogruppo R1b in Italia*. Wikimedia Commons. **March 18, 2009**. https://upload.wikimedia.org/wikipedia/it/archive/d/da/20090720180158!R1bItalia.png

[10] Wikimedia Commons contributors. *Map of total reported cases per capita of the novel coronavirus (COVID-19)*. Wikimedia Commons. **April 27, 2020**. https://upload.wikimedia.org/wikipedia/commons/archive/3/3b/20200427085317!COVID-19_Outbreak_World_Map_per_Capita.svg

[11] Wikimedia Commons contributors. *Haplogroup R1b (Y-DNA)*. Wikimedia Commons. **July 28, 2019**. https://upload.wikimedia.org/wikipedia/commons/e/ec/Haplogroup_R1b_(Y-DNA).PNG

[12] Gómez Moreno, Á. *Coronavirus, Population Genetics, and Humanities*. Mirabilia Journal. Eletronic Journal of Antiquity & Middle Ages, 30. **April 22, 2020**. https://www.revistamirabilia.com/sites/default/files/pdfs/01._gomezmoreno.pdf

[13] Gómez Moreno, Á. *Coronavirus and Genetics: in no way a miracle*. Mirabilia Journal. Eletronic Journal of Antiquity & Middle Ages, 30. **May 5, 2020**. https://www.revistamirabilia.com/sites/default/files/pdfs/02_gomezmoreno.pdf

[14] European Centre for Disease Prevention and Control. *COVID-19 situation update worldwide, as of 29 July 2020*. ECDC. **July 29, 2020**. https://www.ecdc.europa.eu/en/publications-data/download-todays-data-geographic-distribution-covid-19-cases-worldwide

[15] Global Health 50/50. *COVID-19 sex-disaggregated data tracker*. GH5050. **July 24, 2020**. https://globalhealth5050.org/covid19/sex-disaggregated-data-tracker/

[16] Hay, M. *Distribution of European Y-chromosome DNA (Y-DNA) haplogroups by country in percentage*. Eupedia. **June 2017**. http://www.eupedia.com/europe/european_y-dna_haplogroups.shtml

[17] Bentrem, F. W. *COVID-19 Death Rate: Is it in our DNA? (Preprint)*. ResearchGate. **July 17, 2020**. https://doi.org/10.13140/RG.2.2.29960.65289/1

[18] Notari, A., and Torrieri, G. *COVID-19 transmission risk factors (Preprint)*. ArXiv:2005.03651 [Physics, q-Bio, Stat]. **May 7, 2020**. http://arxiv.org/abs/2005.03651

[19] Murray, J. D. *Mathematical Biology: I. An Introduction (3rd edition)*. Springer. **December 8, 2007**. https://books.google.it/books?id=fZYMBwAAQBAJ

[20] Hay, M. *Y-chromosomal haplogroups of the Italians by province and region*. Eupedia. **2013**. https://www.eupedia.com/genetics/regional_italian_y-dna_haplogroups.shtml

[21] Wikipedia contributors. *Genetic history of the Iberian Peninsula*. Wikipedia. **April 15, 2020**. https://en.wikipedia.org/w/index.php?title=Genetic_history_of_the_Iberian_Peninsula&oldid=951003500

[22] Wikipedia contributors. *Y-DNA haplogroups in populations of the Near East*. Wikipedia. **April 18, 2020**. https://en.wikipedia.org/w/index.php?title=Y-DNA_haplogroups_in_populations_of_the_Near_East&oldid=951641460

[23] Korber, B., Fischer, W. M., Gnanakaran, S., et al. *Spike mutation pipeline reveals the emergence of a more transmissible form of SARS-CoV-2 (Preprint)*. BioRxiv, 2020.04.29.069054. **May 5, 2020**. https://doi.org/10.1101/2020.04.29.069054

[24] *Countries in the world by population (2020)*. Worldometer. **2020**. https://www.worldometers.info/world-population/population-by-country/

[25] Wikipedia contributors. *Demographics of Kosovo*. Wikipedia. **June 6, 2020**. https://en.wikipedia.org/w/index.php?title=Demographics_of_Kosovo&oldid=961096947

[26] *Case fatality rate of the ongoing COVID-19 pandemic*. Our World in Data. **July 17, 2020**. https://ourworldindata.org/grapher/coronavirus-cfr

[27] Allen, N., and Khalip, A. *Spain revises coronavirus death toll down by nearly 2,000*. Reuters. **May 25, 2020**. https://www.reuters.com/article/us-health-coronavirus-spain-tally-idUSKBN2311LD

[28] McMurtry, A. *Spain updates COVID-19 death toll to 28,315*. Anadolu Agency. **June 19, 2020**. https://www.aa.com.tr/en/europe/spain-updates-covid-19-death-toll-to-28-315-/1883268

[29] Huheey, J. E., and Martin, D. L. *Malaria, favism and glucose-6-phosphate dehydrogenase deficiency*. Experientia, 31(10), 1145–1147. **October 15, 1975**. https://doi.org/10.1007/BF02326760

[30] Fu, Q., Posth, C., Hajdinjak, M., et al. *The genetic history of Ice Age Europe*. Nature, 534(7606), 200–205. **May 2, 2016**. https://doi.org/10.1038/nature17993

[31] Jin, J.-M., Bai, P., He, W., et al. *Gender Differences in Patients With COVID-19: Focus on Severity and Mortality*. Frontiers in Public Health, 8. **April 29, 2020**. https://doi.org/10.3389/fpubh.2020.00152

[32] *X-linked Recessive: Red-Green Color Blindness, Hemophilia A*. Stanford Children's Health. **2020**. https://www.stanfordchildrens.org/en/topic/default?id=x-linked-recessive-red-green-color-blindness-hemophilia-a-90-P02164

[33] Gemmati, D., Bramanti, B., Serino, M. L., et al. *COVID-19 and Individual Genetic Susceptibility/Receptivity: Role of ACE1/ACE2 Genes, Immunity, Inflammation and Coagulation. Might the Double X-Chromosome in Females Be Protective against SARS-CoV-2 Compared to the Single X-Chromosome in Males?*. International Journal of Molecular Sciences, 21(10), 3474. **May 14, 2020**. https://doi.org/10.3390/ijms21103474

[34] Mutti, L., Pentimalli, F., Baglio, G., et al. *Coronavirus Disease (Covid-19): What Are We Learning in a Country With High Mortality Rate?*. Frontiers in Immunology, 11. **May 28, 2020**. https://doi.org/10.3389/fimmu.2020.01208

[35] Correale, P., Mutti, L., Pentimalli, F., et al. *HLA-B*44 and C*01 Prevalence Correlates with Covid19 Spreading across Italy*. International Journal of Molecular Sciences, 21(15), 5205. **July 23, 2020**. https://doi.org/10.3390/ijms21155205

[36] Ellinghaus, D., Degenhardt, F., Bujanda, L., et al. *Genomewide Association Study of Severe Covid-19 with Respiratory Failure*. New England Journal of Medicine. **June 17, 2020**. https://doi.org/10.1056/NEJMoa2020283

[37] Davies, N. G., Klepac, P., Liu, Y., et al. *Age-dependent effects in the transmission and control of COVID-19 epidemics*. Nature Medicine, 1–7. **June 16, 2020**. https://doi.org/10.1038/s41591-020-0962-9

[38] Poletti, P., Tirani, M., Cereda, D., et al. *Probability of symptoms and critical disease after SARS-CoV-2 infection (Preprint)*. ArXiv:2006.08471 [q-Bio]. **June 15, 2020**. http://arxiv.org/abs/2006.08471

[39] Giangreco, G. *Case fatality rate analysis of Italian COVID-19 outbreak*. Journal of Medical Virology, 92(7), 919–923. **April 16, 2020**. https://doi.org/10.1002/jmv.25894

[40] Heid, M. *Could the Coronavirus Be Weakening as It Spreads?*. Medium. **June 4, 2020**. https://elemental.medium.com/could-the-coronavirus-be-weakening-as-it-spreads-928f2ad33f89

[41] Valenti, L., Bergna, A., Pelusi, S., et al. *SARS-CoV-2 seroprevalence trends in healthy blood donors during the COVID-19 Milan outbreak (Preprint)*. MedRxiv, 2020.05.11.20098442. **May 31, 2020**. https://doi.org/10.1101/2020.05.11.20098442

[42] AFP/The Local. *Coronavirus was already in Italy by December, waste water study shows*. The Local. **June 19, 2020**. https://www.thelocal.it/20200619/coronavirus-was-already-in-italy-by-december-waste-water-study-shows

[43] Lavezzo, E., Franchin, E., Ciavarella, C., et al. *Suppression of a SARS-CoV-2 outbreak in the Italian municipality of Vo'*. Nature. **June 30, 2020**. https://doi.org/10.1038/s41586-020-2488-1

[44] John, T. *Iceland lab's testing suggests 50% of coronavirus cases have no symptoms*. CNN. **April 3, 2020**. https://www.cnn.com/2020/04/01/europe/iceland-testing-coronavirus-intl/index.html

[45] McGraw, E. *A few superspreaders transmit the majority of coronavirus cases*. The Conversation. **June 5, 2020**. http://theconversation.com/a-few-superspreaders-transmit-the-majority-of-coronavirus-cases-139950

[46] Liu, Y., Yan, L.-M., Wan, L., et al. *Viral dynamics in mild and severe cases of COVID-19*. The Lancet. Infectious Diseases, 20(6), 656–657. **March 19, 2020**. https://doi.org/10.1016/S1473-3099(20)30232-2

[47] Booth, R., and Barr, C. *Black people four times more likely to die from Covid-19, ONS finds*. The Guardian. **May 7, 2020**. https://www.theguardian.com/world/2020/may/07/black-people-four-times-more-likely-to-die-from-covid-19-ons-finds

[48] Karson, K., and Scanlan, Q. *Black Americans and Latinos nearly 3 times as likely to know someone who died of COVID-19: POLL*. ABC News. **May 22, 2020**. https://abcnews.go.com/Politics/black-americans-latinos-times-died-covid-19-poll/story?id=70794789

[49] Kloc, M., Ghobrial, R. M., and Kubiak, J. Z. *The Role of Genetic Sex and Mitochondria in Response to COVID-19 Infection*. International Archives of Allergy and Immunology, 181(8), 629–634. **June 19, 2020**. https://doi.org/10.1159/000508560

[50] Maan, A. A., Eales, J., Akbarov, A., et al. *The Y chromosome: a blueprint for men's health?*. European Journal of Human Genetics, 25(11), 1181–1188. **August 30, 2017**. https://doi.org/10.1038/ejhg.2017.128