Title: Person-to-person contagion and antibiotic usage are the key factors responsible for the positive correlation between antibiotic resistance gene diversity and virulence gene diversity in human metagenomes

Célia P. F. Domingues[1] *, João S. Rebelo[1] *, Teresa Nogueira[1, 2], Joël Pothier[3], Francisca Monteiro[1], and Francisco Dionisio[1]

[1] cE3c – Centre for Ecology, Evolution and Environmental Changes, Faculdade de Ciências, Universidade de Lisboa, 1749-016 Lisboa, Portugal

[2] INIAV, Instituto Nacional de Investigação Agrária e Veterinária, I.P. , 2780 -157 Oeiras, Portugal

[3] Atelier de Bioinformatique, ISYEB, UMR 7205 CNRS MNHN UPMC EPHE, Muséum National d'Histoire Naturelle, CP 50, 45 rue Buffon F-75005 Paris, France

* Co-first authors

Corresponding author: Francisco Dionisio

cE3c – Centro de Ecologia, Evolução e Alterações Ambientais, Faculdade de Ciências, Universidade de Lisboa, 1749-016 Lisboa, Portugal

Fax. +351 21750002

dionisio@fc.ul.pt

Keywords: antibiotic resistance; virulence; microbiome; metagenomics; human gut; computer simulation

Abstract

Antibiotics have been used on a large scale in veterinary and medical contexts. The massive and continuous exposure to antimicrobial drugs has led to the emergence and spread of antibiotic resistance, which is nowadays one of the greatest threats to public health. This study aims to understand the cause of the recently shown positive correlation between the diversity of antibiotic resistance genes and the diversity of pathogenicity genes across human metagenomes from different geographical origins. Using computer simulations, we studied the effect of several factors, such as the total diversity of virulence genes and antibiotic resistance genes, the contagion probability between individuals, the loss rate of resistance genes, whether or not the population uses antibiotics, or the contacts' network between people (regular, small-world, and random networks). Our results show that the key factor for the emergence of the positive correlation is the ratio between the contagion probability between individuals and the loss rate of resistance genes. Importantly, if the contagion probability is higher than the loss rate of resistance genes, the positive correlation emerges. We argue that the loss rate of resistance genes may be very low, or even null, leaving contagion probability between people as the main cause for the establishment of a positive correlation between the diversity of antibiotic resistance genes and the diversity of pathogenicity genes across human metagenomes. Therefore, something as prosaic as the contagion between people implies that those humans with higher diversity of virulence genes in their metagenomes are, precisely, those expected to have a high diversity of resistance genes. This is unfortunate because these people with higher diversity of virulence genes are precisely those more plausible to take antibiotics, so the drug is unlikely to be useful.

1. Introduction

Since the 1940's antibiotics have been used, not only in health contexts in veterinary and medicine but also in livestock and agriculture as growth promoters (Castanon, 2007). As an incredible example of Darwinian selection, bacteria around the world have gradually become resistant to several antibiotics. Such spread of resistance has had terrible consequences. For example, there were about 875 500 disability-adjusted life-years and more than 33 000 deaths in European Economic Area due to antibiotic resistance in 2015 (Cassini et al., 2019).

Bacterial communities are often very complex, eventually comprising both pathogenic and non-pathogenic bacteria. The human microbiome, defined as the set of microorganisms that colonize humans (body's surfaces and biofluids, including tissues such as skin, mucosa, and, most importantly, the gastrointestinal tract) is composed of about 3.8×10^{13} bacterial cells (Sender et al., 2016) spanning thousands of taxa. Therefore, even when the objective of uptaking antibiotics is to inhibit growth or kill virulent bacteria, many non-pathogenic (mutualistic or commensal) strains and species are certainly affected. Most likely, the same happens when antibiotics are used as growth promoters.

Pathogenic bacteria can antagonize the host's immune response or alter cell physiology and tissue function (Hornef et al., 2002;Gama et al., 2012). To achieve this goal, they rely on the existence of pathogenic mechanisms or strategies for coordinating the action of virulence factors. Bacterial virulence factors are usually proteins that allow pathogenic bacteria to bind and adhere to host cell targets, invade cells and tissues, exert a toxic effect, acquire iron from host tissues (Chen et al., 2012) or trigger an immune response such as inflammation (Wilson et al., 2002).

As such, we can argue that to colonize a new biome, the bacterial pathogen needs to encode and express a set of virulence factors to have an active pathogenic mechanism. For example, *Staphylococcus aureus* is responsible for skin and soft tissue infections in humans and encodes many different virulence factors. Some of them are unique to the pathogenesis of skin abscesses and contribute to the development of this specific infection: IsdAB for protection against reactive oxygen species; Coa and vWbp

as factors of immune evasion; ClfAB, SdrCDE, FnBPAB, Eap, and Emp, as ECM binding molecules; IsdA for resistance to antimicrobial peptides; Spa and IsdBH for antibody binding and altering phagocytosis; Hla for Cytolysis (Kobayashi et al., 2015). This means that the presence of most of these 13 virulence molecules is an indication of the presence of an abscess and symptoms of skin infection by *S. aureus*.

Exotoxins, such as hemolysins, coagulases, hyaluronidase, or collagenases, are bacterial virulence factors that are also public goods. Its release into the extracellular milieu can benefit any surrounding pathogenic bacteria, in a social cooperative way (Nogueira et al., 2009). The hydrolysis of hemoglobin to release iron, the dissolution of tissues to allow bacterial progression or the chelation of iron, are some examples of pathogenic mechanisms that are common to various infections (Menestrina et al., 2001). Thus, we will group virulence factors into families by type of action, regardless of belonging to the same species/strain.

Symptoms of infection by pathogenic bacteria drive people to seek medical help and take antibiotics, following a pattern of cause and consequence (David et al., 2019). As the selective pressure for antibiotic resistance among pathogens is actively exercised only when there is a disease, one might expect that both resistance and pathogenicity phenotypes are associated. The same could happen at the genotypic level, namely that this pattern of cause and consequence would lead to a positive correlation between the diversity of genes encoding antibiotic resistance and virulence genes.

Indeed, it was recently found that there is a positive correlation between the diversity of antibiotic resistance genes and that of virulence genes across human gut microbiomes (Escudeiro et al., 2019). However, although the exposure to antibiotics in therapeutic doses aims to target pathogens, any other strains or species resistant to the antibiotic also present in the microbiome, including mutualists and commensals, are selected. Moreover, antibiotics are also used in healthy animals as growth promoters, selecting antibiotic-resistance genes among non-pathogens. Therefore, an explanation for the positive correlation mentioned above is still unclear.

Both virulence and resistance genes spread between people and their metagenomes. This may contribute to the accumulation of virulence and resistance genes in some people when they or their contacts take antibiotics. Therefore, contagion between people should play a role in the correlation between the diversity of resistance

and virulence genes. The transmission of microbes from mother to child is already well documented (Blaser and Falkow, 2009;Nayfach et al., 2016;Ferretti et al., 2018;Yassour et al., 2018;Nogueira et al., 2019). A recent study highlighted that the oral and gut microbiomes of individuals belonging to the same household share similarities in bacterial strains, regardless of the individuals' genetic relationship (Brito et al., 2019). This suggests that bacteria in human microbiomes can either have a shared exposure or that they can be transferred from person to person on the social network. This suggestion is supported by a study that showed that social interactions shape the chimpanzee's microbiomes (Moeller et al., 2016).

In this work, we aimed at finding the key factors leading to the positive correlation between the diversity of virulence and antibiotic resistance genes observed across human metagenomes (Escudeiro et al., 2019). To this end, we simulated the transfer of bacterial antibiotic resistance genes and virulence factor genes in a human-to-human transmission network. We show that a positive correlation between the diversity of antibiotic resistance coding genes and those coding for virulence emerges whenever the contagion rate between individuals is higher than the probability that metagenomes lose resistant genes, independently of all the other parameters of the simulations. This simple rule is the major factor explaining the positive correlation between the diversity of virulence genes and the diversity of antibiotic resistance genes.

.

2. Methods

2.1 Building the human network

We simulated a network of people and their respective metagenomes, where each node represents a person's metagenome. To simplify language, sometimes we use the words person or people, meaning a person's metagenome or people's metagenomes, respectively. In the network, the edges represent possible transmission avenues of virulence, resistance, and "other" genes (see below) between two people.

The networks of human contacts were built on a regular network basis, following the methodology of Watts and Strogatz (Watts and Strogatz, 1998). In this case, each node, representing a human being, is linked to the n nearest nodes. To build the network, each link on each node has a certain probability p of being reconnected to another randomly chosen node. The parameter p represents the probability of each connection to be modified. The type of network is defined by the value assigned to the parameter p (for example, a regular network is obtained for $p = 0$, whereas $p = 1$ results in a random network). Unless noted, we performed simulations with $p = 0.5$.

2.2 The metagenome, the immune system and antibiotic uptake

The model considers three types of genes belonging to the human metagenome, namely, antibiotic resistance genes, virulence genes, and the third category of genes that we call "other" genes. This last category represents genes that do not confer resistance to antibiotics nor code for virulence - the "other" genes work as a control to help us to understand to what extent a certain correlation occurs only between virulence genes and antibiotic resistance genes. None of these three gene types (virulence genes, antibiotic resistance genes, and "other" genes) are housekeeping genes.

We focused on the presence (1) or absence (0) of genes encoding for different functions, irrespectively of its copy-number in the metagenome. In the simulations, each gene represents a gene family (with similar functions). Also, we clustered virulence genes into groups of related families, each group having the same number of families. Resistance genes were also divided into groups, each group having the same number of families. Each group represents genes associated with the resistance to an antibiotic. Of

note, we did not consider resistance to multiple drugs in our simulations. Therefore, there will be as many groups as there are antibiotics accounted for in the simulations. The "other" genes are not divided into groups. In this paper, we define the diversity of a certain gene type as the number of genes of that type present in a human metagenome.

During each cycle of the simulation, there is contagion between people (nodes of the network), where virulence, resistance, and "other" genes are transferred with a certain probability. A group of virulence genes represents a complete set of genes that encode a pathogenic mechanism of infection. When, through contagion, one of these groups of virulence genes is complete in a given person, the immune system acts and that person feels ill. In the simulations, the immune system has two different effects: (i) virulence genes of the completed group disappear from the individual metagenome with a given probability; (ii) according to the number of virulence genes of a completed group, the same number of both resistance and "other" genes are randomly chosen and eventually eliminated with the same given probability.

People that feel ill will take antibiotics. Bacteria carrying resistance genes to the antibiotic administered will be selected, whereas the remaining bacteria will be eliminated. In this work, we assumed that all families of resistance genes are present in all metagenomes, but in two different possible states: in some metagenomes, they are present in low concentrations, so they are not transmissible to other individuals in the network; in other metagenomes, these genes were already selected by an antibiotic, so they are transmissible (presumably because their concentration became high).

Moreover, upon antibiotic consumption, the following events occur: (i) resistance genes belonging to the group associated to that antibiotic are selected in the metagenome, meaning that they are present in such high concentrations that they become transferable (ii) resistance genes associated with other antibiotics are lost with a given probability (becoming non-transferable but still present in minute concentrations); (iii) virulence genes and "other" genes disappear from the metagenome with a given probability. Since we are only focusing on the effect of antibiotics and not on their mechanisms of action, all antibiotics considered here can be used to treat ill individuals. Thus, the administered antibiotics will be randomly chosen.

Several processes lead to gene loss. Genes are lost because of the immune system acting on the bacterial pathogens, because of the selective pressure by antibiotics, and because we assume that resistance determinants impose a fitness cost. To take this cost into account, we consider that, in the cases of individuals that do not take antibiotics, a certain percentage of resistance genes is lost in each cycle according to a "loss rate" (with this process, a resistance gene that was transferable to other people, becomes non-transferable).

2.3 Flowchart of the program

Each simulation is composed of several cycles. In each cycle, all procedures described in the pseudocode (Table 1; see also the flowchart in Fig. 1) were considered. We performed exploratory simulations to parameterize our model. A set of parameters was set as default (Table 2). The main steps of the program in each cycle are:

i) Gene transfer (of all three types) between people (i.e. between each node), according to a given contagion probability. With this process, the diversity of genes present in the recipient metagenome increases.

ii) To look for all metagenomes that have, at least, a complete group of virulence genes. In those people the immune system acts, eliminating virulence, resistance and "other" genes according to the immune system curing probability. These people take antibiotics (chosen randomly from a set of possible antibiotics). Because of the antibiotic, the respective resistance genes associated with the chosen antibiotic are selected. Resistance genes not associated with the chosen antibiotic, virulence genes, and "other" genes, are eliminated according to a certain probability. Finally, resistance genes not associated with the chosen antibiotic are eliminated according to the loss rate probability.

iii) The metagenomes of people that did not take antibiotics in this cycle lose resistance genes according to the loss rate probability. This is because we assume that resistance genes impose a fitness cost to their hosts, something not happening with virulence nor "other" genes.

2.4 Statistical analysis

To know if the diversity of antibiotic resistance genes is correlated with the diversity of virulence genes, we considered that Y (diversity of resistance genes) correlates with X (diversity of virulence genes), according to:

$Y = a.X + b$

In this equation, a represents the slope of the linear regression, while b represents the point at which the line crosses the y-axis.

Depending on the context, the null hypothesis is that there is no correlation between the diversity of antibiotic resistance genes and the diversity of virulence genes, or between the diversity of antibiotic resistance genes and the diversity of "other" genes. The alternative hypothesis is that there is a correlation between the diversity of antibiotic resistance genes and the virulence genes or the diversity of "other" genes, respectively. We define $\alpha = 1\times10^{-6}$, rejecting the null hypothesis if P-value $< \alpha$.

Given the complexity of human interactions, it is paramount to simplify the computer simulations. A simplified model allows us to comprehend the effect of specific factors in our simulations, which would otherwise be extremely difficult to detect. As these simplifications do not allow us to make quantitative inferences, we make qualitative analyses. In particular, the focus of the analyses is always on the sign of the correlation (or of the slope of the linear regression) between the diversity of virulence and antibiotic resistance genes (and whether the correlation is significantly different from zero).

Data analyses described above and Student t-tests performed in Supplementary Information were performed in R version 3.5.1 (R Core Team, 2015).

3. Results

3.1 The number of diseases and the likelihood of contagion

This work aims to understand the positive correlation between the diversity of antibiotic resistance genes and virulence genes across human populations observed by Escudeiro et al. (2019). As explained in the Methods section, we assumed that people are organized in a fixed network of contacts and that there is the transmission of genes between connected people. Whenever a group of virulence genes is completed in a person's metagenome (representing a pathogenic mechanism necessary to trigger an infection), the immune system is activated and that person takes an antibiotic. The antibiotic eliminates part of the virulence genes (by eliminating the bacteria encoding them, and for which the antibiotic is targeted to) as well a certain percentage of other genes not associated with the antibiotic administered.

In principle, the contagion probability could have any value in the simulations. Given the importance of the contagion probability, this value must be calibrated according to the other conditions of the model. For that, we assumed that individuals are not affected by more than two infectious diseases at the same time. Therefore, we started this study by searching parameters that lead individuals to have a maximum of two diseases, that is, no more than two full groups of virulence genes at a given cycle.

We performed simulations with different probabilities of contagion and counted the number of complete groups that each individual has per cycle. As we can see in Table 3, when the contagion probability is 0.0025, there is an individual in a certain cycle (out of eight million possibilities) with three groups of virulence genes completed. For this reason, in our simulations, we settled the contagion probability to be 0.002 or less.

3.2 Effect of the relative values of the contagion probability and the resistance gene loss rate

We tested the effect of different combinations of contagion probability and resistance gene loss rate. For that, we fixed all the other parameters (see Table 2). The results obtained after 8000 cycles are shown in Supp. Table1. Fig.2 shows that, if the contagion probability is higher, the same or only slightly lower than the loss rate, the

correlation between the diversity of virulence genes and the diversity of resistance genes is positive, as well as the correlation between the diversity of "other" genes and the diversity of resistance genes (Supp. Table 1, Fig. 2).

3.3 The effect of antibiotic uptake

This section aims to understand the role of antibiotic consumption by the population in the arising of the positive correlation between the diversity of resistance and virulence genes. For that, we performed simulations assuming that no one consumes antibiotics, even if groups of virulence genes are complete in the metagenome.

The results of these simulations are shown in Figure 3A. Even without the use of antibiotics, and with loss rate = 0.0005, there is a correlation between the diversity of resistance genes and the diversity of virulence genes, and also with the diversity of "other" genes (virulence genes: R = 0.360, slope = 0.769 , P-value = 5.05×10^{-32}; "other" genes: R = 0.428, slope = 0.631, P-value = 7.22×10^{-46}). However, there is an important difference between Figs 2A and 3A. In Fig. 3A, the diversity of both the virulence and the "other" genes tend to reach high values in each person (all individuals have approximately 60 or more of those genes). Such a density of points with a high diversity of genes is not seen when antibiotics are taken (Figure 2A) nor in (real) human populations (Escudeiro et al., 2019).

In Fig. 3A, we can also see that the diversity of virulence and "other" genes is higher than that of resistance genes. However, if there is no loss rate (loss rate = 0), the diversity of resistance genes reaches higher values (Fig. 3B).

3.4 Strength of the positive correlation if the contagion probability is higher than the transmission rate.

We have seen that the positive correlation between the diversity of virulence and resistance genes is positive if the contagion probability is higher than the loss rate. In this section we check the robustness of this result. For that, in the next five subsections, we studied the impact of changing the parameters of the simulations, or changing the network itself. The parameters studied are: population size, the proportion

between virulence/"other" genes and antibiotic resistance genes, the elimination probability under antibiotic intake, the proportion of the population harboring antibiotic-resistance genes in their metagenome, and the network type.

3.4.1 Impact of the population size

We performed simulations with a higher population size, namely 5000 individuals for nine of the 24 conditions shown in Fig. 2C, namely: the nine possible combinations of loss rate = {0.0005, 0.0015, 0.0025} and contagion probability = {0.0005, 0.0010, 0.0015}. The slopes are not significantly different from the cases where the population size was 1 000 individuals (Supp. Table 2). An increase in the population size leads to an increase in the number of intermediaries between two distant individuals. Therefore, for virulence genes and antibiotic resistance genes to be transferred between these two distant individuals, more contacts are needed and, consequently, more time is required to achieve a stable correlation. Importantly, the final correlation does not vary with the population size (Supp. Table 2).

3.4.2 Impact of different proportions between virulence/ "other" genes and antibiotic resistance genes.

In all the other sections, we consider equal diversity of all types of genes. Here, we studied the effect of considering that the diversity of virulence and "other" genes is different from the diversity of resistance genes for the same nine conditions of contagion probability and loss rate as in the previous section. Although there were significant differences between the slopes in 29 out of 36 cases, we didn't observe a change in the sign of the correlation slope (Supp. Table 3).

3.4.3 Impact of the antibiotics

In all other sections, when an antibiotic is taken, virulence genes, "other" genes, and resistance genes are eliminated from the metagenome with a probability of 0.7

(except for resistance genes corresponding to the antibiotic used, which are selected). In this section, we analyzed the impact of using other elimination probabilities. For that, we performed simulations similar to the previous ones, for the same nine conditions of contagion probability and loss rate, but where the probability of eliminating genes under antibiotic intake is 0.3 and 0.5 for all gene types (instead of 0.7). In 13 out of 18 cases, the slopes were not significantly different from the ones obtained with a probability of 0.7 (Supp. Table 4). In the other five cases, the slopes are different but the sign remains (Supp. Table 4).

We also checked the impact of setting the probability of eliminating antibiotic resistance genes different from that of eliminating virulence and "other" genes. Although the slopes were significantly different in six out of 54 tested cases (Suppl. Table 5), the sign of the slopes remained. Overall, these results show that the sign of the slope is robust under changes in the elimination probability.

3.4.4 Impact of the initial proportion of metagenomes containing antibiotic-resistance genes

In the other sections we considered that initially in the simulation, all individuals carry all the antibiotic resistance genes in two alternative states: either unselected, being unable to be transmitted to other people because they are present at low concentrations, or they were previously selected with an antibiotic, hence being able to pass to other people. In this section, we study the effect of considering that, initially, only 10% of the metagenomes contain antibiotic-resistance genes. With this parameter changed, the simulations take more time to stabilize because 90% of the population receives resistance genes only through contagion. We performed a simulation similar to the ones shown in Figure 2A, but with 20000 cycles. The final slope is not significantly different from the case where all metagenomes initially contain antibiotic-resistance genes (Supp. Table 6).

3.4.5 Impact of the network type

The simulations leading to Figure 2 were performed in a network with a rewiring probability of p = 0.5 (see Methods). We then performed similar simulations but in a regular (p = 0) and in a random (p = 1) network, for the four conditions of loss rate = {0.0005, 0.0015} and contagion probability = {0.0005, 0.0015}. With p = 0.5 or p = 1, any pair of individuals have a small average shortest path length. This is not the case of p = 0, where the average shortest path length is maximized. Therefore, the number of cycles for the stabilization of simulations was much higher for p = 0 than in the cases of p = 0.5 or p = 1 (Supp. Fig. 1). Accordingly, for p = 0, the program had to run 60 000 cycles for the simulations to stabilize, instead of 8 000 cycles used for p = 0.5 and p = 1. Despite these differences in the time necessary for stabilization, the final slopes are not significantly different from the ones obtained with p = 0.5 (P-value > 1x10^{-6}) (Supp. Table 7; Supp. Fig. 1).

4. Discussion

Even when antibiotics target specific bacterial pathogens, hundreds of other bacterial strains and species are affected. Moreover, antibiotics are frequently given to healthy animals and plants as growth-promoters. With these two processes, the frequency of antibiotic resistance raises in metagenomes. Meanwhile, metagenomes from sick or healthy people harbor many virulence genes. While these facts, together with the contagion of microbiomes between people, surely contribute to the positive correlation between the diversity of virulence and antibiotic-resistance genes observed among human populations (Escudeiro et al., 2019), its causes are still unclear. The objective of this work was to understand what causes this positive correlation.

The main result of our simulations is that a positive correlation emerges if the contagion probability is higher than the loss rate of antibiotic-resistance genes. This can be understood in the following way. In each cycle, some individuals do not need to take an antibiotic in that cycle, because there is no complete group of virulence genes in their metagenome. In these individuals, the diversity of both virulence and "other" genes increase through contagion with other people. However, the diversity of resistance genes only increases if the effect of contagion is stronger than that of gene loss. Therefore, if the contagion rate is higher than the loss rate, a positive correlation emerges between the diversity of antibiotic resistance genes and virulence genes, as well as between resistance and "other" genes.

At first, one might expect to observe a negative correlation whenever the contagion probability was lower than the loss rate. This is not the case when the contagion probability was only slightly lower than the loss rate. For example, if the contagion probability is 0.0005 and the loss rate is 0.0010 the correlation is still positive (Fig. 2, Supp. Table 1). The reason for this is that, in each cycle, one individual contact four other individuals, and, during each of these contacts, they share bacteria from its microbiomes. Each individual, in turn, can only be medicated with antibiotics once (at the end of a cycle), which implies that the rate of loss of resistance genes applies only once in a cycle. Therefore, the impact of the contagion rate is higher than the rate of loss of resistance genes.

Our result shows that a positive correlation emerges if the contagion probability is higher than the loss rate of antibiotic-resistance genes, which is robust under changes of the population size (Supp. Table 2), the network type (Supp. Table 7), the proportion between virulence/"other" genes and antibiotic resistance genes (Supp. Table 3) or the elimination probability under antibiotic intake (Supp. Tables 4 and 5).

We originally assumed that resistance determinants are already present in tiny amounts in every human metagenome because they have been selected through the massive usage of antibiotics since the 1940s. What is the impact of this assumption? As shown in Supp. Table 6, if we, instead, assume that initially, only 10% of the metagenomes contain antibiotic-resistance genes, the final correlations between the diversity of resistance genes and the diversity of virulence or "other" genes are the same. The only difference is that more cycles are needed to stabilize the correlation

We also studied what happens if no one in the population takes antibiotics. Fig. 3A shows that, at the end of simulations, there is a wider range of resistance genes diversity than of the other two groups of genes (it can be seen that the diversity of the virulence and "other" genes are closer to the maximum in Fig. 3A than in Fig. 2A). The cause for such difference between the diversity of resistance genes, and the diversity of the other two gene types, is that resistance genes have a certain probability of being lost as a consequence of a putative fitness cost of resistance determinants. If the loss rate is zero, the diversity of resistance genes becomes higher in all metagenomes (Fig. 3B) than when there is a fitness cost (loss rate higher than zero, Fig. 3A). Note that the rise of the diversity of virulence genes is transient because, as soon as one of the groups of virulence genes is completed, the immune system is activated to eliminate this entire group of virulence genes, as well as some genes from the other two gene types. Therefore, the diversity of "other" genes reaches slightly higher values than that of virulence genes (Fig. 3A and 3B).

Importantly, when we compare our results in Fig. 3 (without antibiotic consumption) with those obtained in Fig. 2A (with antibiotic consumption), we realize that the simulations that best support Escudeiro et al (2019)'s results obtained from real human populations are the ones where sick individuals take antibiotics (Fig. 2A) - in this case, the diversity of virulence and resistance genes in human metagenomes are not clustered at high values (Escudeiro et al., 2019).

The contagion probability between people and the loss rate of antibiotic-resistance genes are the two important parameters of our main result, so it is relevant to know their actual values. In recent years, the interest in human microbiomes strongly increased, yet we still do not know how much is the contagion probability of non-housekeeping genes such as the ones studied here. We know, for example, that human microbiomes are more similar among humans living together, irrespective of the genetic relatedness, suggesting that transmission is a key factor of the microbiome constitution (Rothschild et al., 2018).

Sarowska and colleagues recently reviewed the fate of the so-called extraintestinal pathogenic *Escherichia coli* (ExPEC), which are facultative pathogens of the normal human intestinal microbiome. Many virulence genes are involved in the pathogenicity of ExPEC and some of them are mostly encoded on pathogenicity islands or mobile genetic elements such as plasmids. One of the authors' conclusions is precisely the difficulty in assigning ExPEC transmission to people due to the delay between ExPEC colonization and infection: ExPEC can live in human intestines for months or even years before starting an infection (Sarowska et al., 2019). The same problem applies to the transmission rate of antibiotic-resistance genes: there is very little data on transmission rates between people (Andersson and Hughes, 2017).

We have seen that the relationship between the contagion rate and loss rate is paramount. How much is the loss rate of resistance determinants in human metagenomes? Several longitudinal studies have shown that antibiotic-resistance genes often remain tens of days, sometimes months in human guts (Horcajada et al., 2002;Lautenbach et al., 2006;Weinreich et al., 2006;O'Fallon et al., 2009;Rogers et al., 2012). While still harboring resistance genes, people most probably contact with several other people. Yet, it is still unclear what is the relationship between contagion and loss rates.

As explained in the methods section, the loss of antibiotic resistance determinants is a consequence of the fitness cost of resistance determinants on bacterial cells (compared to otherwise isogenic susceptible cells). Several studies have shown that resistance determinants, here broadly comprising resistance mutations and resistance genes often encoded in plasmids, impose a fitness cost on their hosts (giving the sensitive strains a growth advantage) (Andersson and Levin, 1999). However, several

mechanisms have been shown to decrease or even eliminate fitness costs. First, compensatory mutations, which mask the deleterious effects of resistance mutations, have been observed in several studies (Levin et al., 1997;Schrag et al., 1997;Bjorkman et al., 2000;Maisnier-Patin and Andersson, 2004;Nilsson et al., 2006). Resistance mutations can even be beneficial in certain resistance genetic backgrounds (Trindade et al., 2009). Third, while resistance plasmids often impose a fitness cost to their hosts, is has also been observed that plasmid and/or cells need just a few hundreds of bacterial generations to adapt to each other (Bouma and Lenski, 1988;Modi and Adams, 1991;Dahlberg and Chao, 2003;Dionisio et al., 2005;San Millan et al., 2014;Harrison et al., 2015). Fourth, plasmids exist that increase the fitness of bacteria already harbouring a resistance mutation (Silva et al., 2011); likewise, some resistance mutations increase the fitness of plasmid bearing cells (Silva et al., 2011). The same may happen with two plasmids: one of them compensating for the cost of the other (Silva et al., 2011;San Millan et al., 2014). Fifth, plasmids may interact with other plasmids in such a way as facilitating their transfer (Gama et al., 2017b;a). Sixth, it has been suggested that most plasmids are perhaps costly if their fitness effect is measured within a short time of its isolation from nature (Lau et al., 2013;Gama et al., 2018).

Together, these six factors suggest that the fitness cost of resistance determinants is often very low or null, allowing the permanence of resistance determinants in microbiomes for long periods. This implies that the loss rate of resistance determinants, that is, the probability that a metagenome loses a certain resistance gene or mutation, is certainly lower than the contagion probability. Therefore it is mostly the contagion between people that leads to the positive correlation between the diversity of resistance genes and virulence genes, or in fact, between the diversity of resistance genes and other gene types not involved in pathogenicity mechanisms.

5. Concluding remarks

The simple fact that people contaminate between themselves, and, of course, the almost omnipresence of resistance determinants in human metagenomes, is chief to explain the positive correlation between antibiotic resistance gene diversity and virulence gene diversity across human metagenomes. This is a strong and general result

because very few assumptions were made. This has worrying health implications: people requiring antibiotics are: (1) most likely, those with higher diversity of virulence genes in their metagenomes, and, (2) unfortunately, those with higher diversity of resistance determinants. Therefore, the chances that the antibiotic will not work due to resistance are high.

Acknowledgments

Teresa Nogueira was supported by contract ALG-01-0145-FEDER-028824 and Francisca Monteiro acknowledges FCT-Fundação para a Ciência e a Tecnologia, I.P., for her Postdoc fellowship grant SFRH/BPD/123504/2016. FCT supports cE3c by contract UIDP/00329/2020.

References

Andersson, D.I., and Hughes, D. (2017). Selection and Transmission of Antibiotic-Resistant Bacteria. *Microbiology Spectrum* 5.

Andersson, D.I., and Levin, B.R. (1999). The biological cost of antibiotic resistance. *Current Opinion in Microbiology* 2**,** 489-493.

Bjorkman, J., Nagaev, I., Berg, O.G., Hughes, D., and Andersson, D.I. (2000). Effects of environment on compensatory mutations to ameliorate costs of antibiotic resistance. *Science* 287**,** 1479-1482.

Blaser, M.J., and Falkow, S. (2009). What are the consequences of the disappearing human microbiota? *Nature Reviews Microbiology* 7**,** 887-894.

Bouma, J.E., and Lenski, R.E. (1988). Evolution of a Bacteria Plasmid Association. *Nature* 335**,** 351-352.

Brito, I.L., Gurry, T., Zhao, S.J., Huang, K., Young, S.K., Shea, T.P., Naisilisilis, W., Jenkins, A.P., Jupiter, S.D., Gevers, D., and Alm, E.J. (2019). Transmission of human-associated microbiota along family and social networks. *Nature Microbiology* 4**,** 964-971.

Cassini, A., Hogberg, L.D., Plachouras, D., Quattrocchi, A., Hoxha, A., Simonsen, G.S., Colomb-Cotinat, M., Kretzschmar, M.E., Devleesschauwer, B., Cecchini, M., Ouakrim, D.A., Oliveira, T.C., Struelens, M.J., Suetens, C., Monnet, D.L., Strauss, R., Mertens, K., Struyf, T., Catry, B., Latour, K., Ivanov, I.N., Dobreva, E.G., Andrasevic, A.T., Soprek, S., Budimir, A., Paphitou, N., Zemlickova, H., Olsen, S.S., Sonksen, U.W., Martin, P., Ivanova, M., Lyytikainen, O., Jalava, J., Coignard, B., Eckmanns, T., Abu Sin, M., Haller, S., Daikos, G.L., Gikas, A., Tsiodras, S., Kontopidou, F., Toth, A., Hajdu, A., Guolaugsson, O., Kristinsson, K.G., Murchan, S., Burns, K., Dsstat, P.P., Gagliotti, C., Dumpis, U., Liuimiene, A., Perrin, M., Borg, M.A., De Greeff, S.C., Monen, J.C.M., Koek, M.B.G., Elstrom, P., Zabicka, D., Deptula, A., Hryniewicz, W., Canica, M., Nogueira, P.J., Fernandes, P.A., Manageiro, V., Popescu, G.A., Serban, R.I., Schreterova, E., Litvova, S., Stefkovicova, M., Kolman, J., Klavs, I., Korosec, A., Aracil, B., Asensio, A., Perez-Vazquez, M., Billstrom, H., Larsson, S., Reilly, J.S., Johnson, A., Hopkins, S., and Grp, B.a.C. (2019). Attributable deaths and disability-adjusted life-years caused by infections with antibiotic-resistant bacteria in the EU and the European Economic Area in 2015: a population-level modelling analysis. *Lancet Infectious Diseases* 19**,** 56-66.

Castanon, J.I.R. (2007). History of the use of antibiotic as growth promoters in European poultry feeds. *Poultry Science* 86**,** 2466-2471.

Chen, L.H., Xiong, Z.H., Sun, L.L., Yang, J., and Jin, Q. (2012). VFDB 2012 update: toward the genetic diversity and molecular evolution of bacterial virulence factors. *Nucleic Acids Research* 40**,** D641-D645.

Dahlberg, C., and Chao, L. (2003). Amelioration of the cost of conjugative plasmid carriage in Eschericha coli K12. *Genetics* 165**,** 1641-1649.

David, P.H.C., Sá-Pinto, X., and Nogueira, T. (2019). Using SimulATe to model the effects of antibiotic selective pressure on the dynamics of pathogenic bacterial populations. *Biology Methods and Protocols* 4**,** bpz004.

Dionisio, F., Conceicao, I.C., Marques, A.C., Fernandes, L., and Gordo, I. (2005). The evolution of a conjugative plasmid and its ability to increase bacterial fitness. *Biol Lett* 1**,** 250-252.

Escudeiro, P., Pothier, J., Dionisio, F., and Nogueira, T. (2019). Antibiotic Resistance Gene Diversity and Virulence Gene Diversity Are Correlated in Human Gut and Environmental Microbiomes. *mSphere* 4.

Ferretti, P., Pasolli, E., Tett, A., Asnicar, F., Gorfer, V., Fedi, S., Armanini, F., Truong, D.T., Manara, S., Zolfo, M., Beghini, F., Bertorelli, R., De Sanctis, V., Bariletti, I., Canto, R., Clementi, R., Cologna, M., Crifo, T., Cusumano, G., Gottardi, S., Innamorati, C., Mase, C., Postai, D., Savoi, D., Duranti, S., Lugli, G.A., Mancabelli, L., Turroni, F., Ferrario, C., Milani, C., Mangifesta, M., Anzalone, R., Viappiani, A., Yassour, M., Vlamakis, H., Xavier, R., Collado, C.M., Koren, O., Tateo, S., Soffiati, M., Pedrotti, A., Ventura, M., Huttenhower, C., Bork, P., and Segata, N. (2018). Mother-to-Infant Microbial Transmission from Different Body Sites Shapes the Developing Infant Gut Microbiome. *Cell Host & Microbe* 24**,** 133-+.

Gama, J.A., Abby, S.S., Vieira-Silva, S., Dionisio, F., and Rocha, E.P.C. (2012). Immune Subversion and Quorum-Sensing Shape the Variation in Infectious Dose among Bacterial Pathogens. *Plos Pathogens* 8.

Gama, J.A., Zilhao, R., and Dionisio, F. (2017a). Conjugation efficiency depends on intra and intercellular interactions between distinct plasmids: Plasmids promote the immigration of other plasmids but repress co-colonizing plasmids. *Plasmid* 93**,** 6-16.

Gama, J.A., Zilhao, R., and Dionisio, F. (2017b). Multiple plasmid interference - Pledging allegiance to my enemy's enemy. *Plasmid* 93**,** 17-23.

Gama, J.A., Zilhao, R., and Dionisio, F. (2018). Impact of plasmid interactions with the chromosome and other plasmids on the spread of antibiotic resistance. *Plasmid*.

Harrison, E., Guymer, D., Spiers, A.J., Paterson, S., and Brockhurst, M.A. (2015). Parallel Compensatory Evolution Stabilizes Plasmids across the Parasitism-Mutualism Continuum. *Current Biology* 25**,** 2034-2039.

Horcajada, J.P., Vila, J., Moreno-Martinez, A., Ruiz, J., Martinez, J.A., Sanchez, M., Soriano, E., and Mensa, J. (2002). Molecular epidemiology and evolution of resistance to quinolones in Escherichia coli after prolonged administration of ciprofloxacin in patients with prostatitis. *Journal of Antimicrobial Chemotherapy* 49**,** 55-59.

Hornef, M.W., Wick, M.J., Rhen, M., and Normark, S. (2002). Bacterial strategies for overcoming host innate and adaptive immune responses. *Nature Immunology* 3**,** 1033-1040.

Kobayashi, S.D., Malachowa, N., and Deleo, F.R. (2015). Pathogenesis of Staphylococcus aureus Abscesses. *American Journal of Pathology* 185**,** 1518-1527.

Lau, B.T.C., Malkus, P., and Paulsson, J. (2013). New quantitative methods for measuring plasmid loss rates reveal unexpected stability. *Plasmid* 70**,** 353-361.

Lautenbach, E., Tolomeo, P., Mao, X.Q., Fishman, N.O., Metlay, J.P., Bilker, W.B., and Nachamkin, I. (2006). Duration of outpatient fecal colonization due to Escherichia coli isolates with decreased susceptibility to fluoroquinolones: Longitudinal study of patients recently discharged from the hospital. *Antimicrobial Agents and Chemotherapy* 50**,** 3939-3943.

Levin, B.R., Lipsitch, M., Perrot, V., Schrag, S., Antia, R., Simonsen, L., Walker, N.M., and Stewart, F.M. (1997). The population genetics of antibiotic resistance. *Clinical Infectious Diseases* 24**,** S9-S16.

Maisnier-Patin, S., and Andersson, D.I. (2004). Adaptation to the deleterious effects of antimicrobial drug resistance mutations by compensatory evolution. *Research in Microbiology* 155**,** 360-369.

Menestrina, G., Serra, M.D., and Prevost, G. (2001). Mode of action of beta-barrel pore-forming toxins of the staphylococcal alpha-hemolysin family. *Toxicon* 39**,** 1661-1672.

Modi, R.I., and Adams, J. (1991). Coevolution in Bacterial-Plasmid Populations. *Evolution* 45**,** 656-667.

Moeller, A.H., Foerster, S., Wilson, M.L., Pusey, A.E., Hahn, B.H., and Ochman, H. (2016). Social behavior shapes the chimpanzee pan-microbiome. *Science Advances* 2.

Nayfach, S., Rodriguez-Mueller, B., Garud, N., and Pollard, K.S. (2016). An integrated metagenomics pipeline for strain profiling reveals novel patterns of bacterial transmission and biogeography. *Genome Research* 26**,** 1612-1625.

Nilsson, A.I., Zorzet, A., Kanth, A., Dahlstrom, S., Berg, O.G., and Andersson, D.I. (2006). Reducing the fitness cost of antibiotic resistance by amplification of initiator tRNA genes. *Proceedings of the National Academy of Sciences of the United States of America* 103**,** 6976-6981.

Nogueira, T., David, P.H.C., and Pothier, J. (2019). Antibiotics as both friends and foes of the human gut microbiome: The microbial community approach. *Drug Development Research* 80**,** 86-97.

Nogueira, T., Rankin, D.J., Touchon, M., Taddei, F., Brown, S.P., and Rocha, E.P.C. (2009). Horizontal Gene Transfer of the Secretome Drives the Evolution of Bacterial Cooperation and Virulence. *Current Biology* 19**,** 1683-1691.

O'fallon, E., Gautam, S., and D'agata, E.M.C. (2009). Colonization with Multidrug-Resistant Gram-Negative Bacteria: Prolonged Duration and Frequent Cocolonization. *Clinical Infectious Diseases* 48**,** 1375-1381.

R Core Team (2015). "R: A Language and Environment for Statistical Computing". (Vienna, Austria: R Foundation for Statistical Computing).

Rogers, B.A., Kennedy, K.J., Sidjabat, H.E., Jones, M., Collignon, P., and Paterson, D.L. (2012). Prolonged carriage of resistant E-coli by returned travellers: clonality, risk factors and bacterial characteristics. *European Journal of Clinical Microbiology & Infectious Diseases* 31**,** 2413-2420.

Rothschild, D., Weissbrod, O., Barkan, E., Kurilshikov, A., Korem, T., Zeevi, D., Costea, P.I., Godneva, A., Kalka, I.N., Bar, N., Shilo, S., Lador, D., Vila, A.V., Zmora, N., Pevsner-Fischer, M., Israeli, D., Kosower, N., Malka, G., Wolf, B.C., Avnit-Sagi, T., Lotan-Pompan, M., Weinberger, A., Halpern, Z., Carmi, S., Fu, J.Y., Wijmenga, C., Zhernakova, A., Elinav, E., and Segal, E. (2018). Environment dominates over host genetics in shaping human gut microbiota. *Nature* 555**,** 210-+.

San Millan, A., Heilbron, K., and Maclean, R.C. (2014). Positive epistasis between co-infecting plasmids promotes plasmid survival in bacterial populations. *Isme Journal* 8**,** 601-612.

Sarowska, J., Futoma-Koloch, B., Jama-Kmiecik, A., Frej-Madrzak, M., Ksiazczyk, M., Bugla-Ploskonska, G., and Choroszy-Krol, I. (2019). Virulence factors, prevalence and potential transmission of extraintestinal pathogenic Escherichia coli isolated from different sources: recent reports. *Gut Pathogens* 11.

Schrag, S.J., Perrot, V., and Levin, B.R. (1997). Adaptation to the fitness costs of antibiotic resistance in Escherichia coli. *Proceedings of the Royal Society B-Biological Sciences* 264**,** 1287-1291.

Sender, R., Fuchs, S., and Milo, R. (2016). Revised Estimates for the Number of Human and Bacteria Cells in the Body. *Plos Biology* 14.

Silva, R.F., Mendonça, S.C., Carvalho, L.M., Reis, A.M., Gordo, I., Trindade, S., and Dionisio, F. (2011). Pervasive sign epistasis between conjugative plasmids and drug-resistance chromosomal mutations. *PLoS genetics* 7**,** e1002181.

Trindade, S., Sousa, A., Xavier, K.B., Dionisio, F., Ferreira, M.G., and Gordo, I. (2009). Positive epistasis drives the acquisition of multidrug resistance. *PLoS genetics* 5**,** e1000578.

Watts, D.J., and Strogatz, S.H. (1998). Collective dynamics of 'small-world' networks. *Nature* 393**,** 440-442.

Weinreich, D.M., Delaney, N.F., Depristo, M.A., and Hartl, D.L. (2006). Darwinian evolution can follow only very few mutational paths to fitter proteins. *Science* 312**,** 111-114.

Wilson, J.W., Schurr, M.J., Leblanc, C.L., Ramamurthy, R., Buchanan, K.L., and Nickerson, C.A. (2002). Mechanisms of bacterial pathogenicity. *Postgraduate Medical Journal* 78**,** 216-224.

Yassour, M., Jason, E., Hogstrom, L.J., Arthur, T.D., Tripathi, S., Siljander, H., Selvenius, J., Oikarinen, S., Hyoty, H., Virtanen, S.M., Ilonen, J., Ferretti, P., Pasolli, E., Tett, A., Asnicar, F., Segata, N., Vlamakis, H., Lander, E.S., Huttenhower, C., Knip, M., and Xavier, R.J. (2018). Strain-Level Analysis of Mother-to-Child Bacterial Transmission during the First Few Months of Life. *Cell Host & Microbe* 24**,** 146-+.

Figure legends

Figure 1. Flowchart of the program. Construction of the network is followed by contact between nodes (people) and, eventually, gene transfer. Some individuals become ill and take antibiotics. Some genes are lost due to the immune system, the antibiotics, and fitness cost of resistance genes.

Figure 2. Effect of the relative values of the contagion probability and the resistance genes loss rate. A and B: the relationship between the diversity of resistance genes and the diversity of the other two gene types. The vertical axes represent the diversity of resistance genes. The horizontal axes represent the diversity of virulence or the diversity of "other" genes. Each blue and orange dot represents the diversity of virulence genes and the "other" genes present in an individual metagenome, respectively. A: contagion probability = 0.001 > resistance genes loss rate = 0.0005, resulting in a positive slope; (virulence genes: R = 0.861, slope = 0.597, p-value = 2.22×10^{-295}; "other" genes**:** R = 0.858, slope = 0.566, p-value = 3.32×10^{-291}) B: contagion probability = 0.001 < resistance genes loss rate = 0.003, resulting in a negative slope; (virulence genes: R = -0.195, slope = -0.065, p-value = 4.62×10^{-10}; "other" genes**:** R = -0.198, slope = -0.063, p-value = 2.84×10^{-10}) C: Slope of the regression between the diversity of virulence genes and resistance genes according to the contagion probability and the loss rate. The horizontal axis represents the values of the resistance genes loss rate. The vertical axis represents the values of the contagion probability. The green color represents the simulations where the slopes are positive; the red color represents the simulations where the slopes are negative; the blue color represents the simulation where the slope is not significantly different from zero (p-value < 1×10^{-6}).

Figure 3. Relationship between resistance genes and virulence or "other" genes when there is no antibiotic use. As in Fig.2, the vertical axes represent the diversity of resistance genes and the horizontal axes represent the diversity of virulence or the diversity of "other" genes. Each blue and orange dot represents the diversity of virulence genes and of the "other" genes present in an individual metagenome, respectively.

Simulations were performed using the parameters described in Methods (Table 2), with contagion probability of 0.001. Virulence and "other" genes accumulate because the only factor that leads to their decrease is the effect of the immune system. A: The diversity of resistance genes also decrease due to the cost of resistance genes (loss rate = 0.0005), thus creating a wider range of resistance genes diversity than the other two groups of genes (virulence genes: R = 0.360, slope = 0.769, p-value = 5.05×10^{-32} ; "other" genes: R = 0.428, slope = 0.631, p-value = 7.22×10^{-46}). B: In this case, there is no loss rate (loss rate = 0), so only the immune system is responsible for the decrease of resistance genes. Therefore, in this case, all gene types accumulate (virulence genes: R = 0.530, slope = 0.780, p-value = 2.28×10^{-73} ; "other" genes: R = 0.645, slope = 0.626, p-value = 1.69×10^{-118}).

Figure 1

Figure 2

A contagion probability = 0.001   B contagion probability = 0.001
and resistance genes loss rate = 0.0005   and resistance genes loss rate = 0.003

Figure 3

Process	Pseudo Code
Table 1. Pseudocode of the program*	
Process	**Pseudo Code**
Gene transfer	For each connection between two individuals do (for each individual of the connection do (get the genes present in each individual metagenome; transmit genes to the other individual of the connection according to the contagion probability))
Screening of individuals	For each individual do (for each group of virulence genes do (check if the group is completed))
Immune system effect	For each group of virulence genes completed do (eliminate virulence genes according to the immune system curing probability). Select as many resistance genes as virulence genes present in the completed groups; eliminate resistance genes according to the immune system curing probability. Select as many "other" genes as virulence genes present in the completed groups; eliminate "other" genes according to the immune system curing probability.
Antibiotic effect	Choose an antibiotic randomly. Select all resistance genes associated with the chosen antibiotic. Eliminate resistance genes not associated with the chosen antibiotic according to the probability of eliminating genes under antibiotic intake Eliminate virulence genes according to the probability of eliminating genes under antibiotic intake. Eliminate "other" genes according to the probability of eliminating genes under antibiotic intake
Loss rate of resistance genes under antibiotic consumption	Eliminate resistance genes not associated with the chosen antibiotic according to the loss rate probability.
Loss rate of resistance genes without antibiotic consumption	Eliminate resistance genes according to the loss rate probability.

*The program code was implemented in the Python programming language.

Table 2. Parameters and default values used in simulations.

Parameters	Default values	Changing values
Rewiring connectivity probability p	0.5	0 or 1
Number of individuals	1000	5000
Number of virulence genes	100	200, 400
Number of resistance genes	100	200, 400
Number of "other" genes	100	200, 400
Number of antibiotics	5	NA
Contagion probability	0.001	0.0005, 0.0015, 0.002
Probability of eliminating genes under the immune system	0.5	NA
Probability of eliminating genes under antibiotic intake	0.7	0.3, 0.5
The loss rate of resistance genes	0.0005	0.001, 0.0015, 0.002, 0.0025, 0.003

Table 3. Full groups of virulence genes according to the contagion probability.

		Number of full groups of virulence genes (in 8 000 000 possibilities)				
		1	**2**	**3**	**4**	**5**
Contagion probability	**0.0005**	3415	2	0	0	0
	0.001	9545	14	0	0	0
	0.0015	15746	34	0	0	0
	0.002	21886	85	0	0	0
	0.0025	28389	146	1	0	0