What is the role of the environment in the emergence of novel antibiotic resistance genes? – A modelling approach

Johan Bengtsson-Palme^{1,2*}, Viktor Jonsson³, Stefanie Heß⁴

¹ Department of Infectious Diseases, Institute of Biomedicine, The Sahlgrenska Academy, University of Gothenburg, Guldhedsgatan 10, SE-413 46, Gothenburg, Sweden

² Centre for Antibiotic Resistance research (CARE) at University of Gothenburg, Gothenburg, Sweden

³ Integrated Science Lab, Department of Physics, Umeå University, SE-901 87 Umeå, Sweden

⁴ Institute of Microbiology, Technische Universität Dresden, Zellescher Weg 20b, 01847 Dresden, Germany

* Corresponding author: joan.bengtsson-palme@microbiology.se

Abstract

It is generally accepted that intervention strategies to curb antibiotic resistance cannot solely focus on human and veterinary medicine but must also consider environmental settings. While the environment clearly has a role in the transmission of resistant bacteria, it is less clear what role it plays in the emergence of novel types of resistance. It has been suggested that the environment constitutes an enormous recruitment ground for resistance genes to pathogens, but the extent to which this actually happens is unknown. In this study, we built a model framework for resistance emergence and used the available quantitative data on the relevant processes to identify the steps which are limiting the appearance of antibiotic resistance determinants in human or animal pathogens. We also assessed the effect of uncertainty in the available data on the model results. We found that in a majority of scenarios, the environment would only play a minor role in the emergence of novel resistance genes. However, the uncertainty around this role is enormous, highlighting an urgent need of more quantitative data to understand the role of the environment in antibiotic resistance development. Specifically, more data is most needed on the fitness costs of antibiotic resistance gene (ARG) carriage, the degree of dispersal of resistant bacteria from the environment to humans, but also the rates of mobilization and horizontal transfer of ARGs. Quantitative data on these processes is instrumental to determine which processes that should be targeted for interventions to curb development and transmission of resistance.

Introduction

Antibiotic resistance is a globally growing health threat, which is projected to take more lives than all forms of cancer combined by 2050 if it cannot be controlled ¹. Nowadays, there is a general consensus that intervention strategies should consider not only human and veterinary medicine but also take into account the environment, in what is referred to as a one health approach ². While surveillance programs and careful reduction of the use of antibiotics in clinical and agricultural settings started years ago, we are still only starting to understand the role of the environment in the development and dissemination of antibiotic resistance ^{3–5}. Based on monitoring data on the abundance and diversity of antibiotic resistance genes (ARGs), we know conceptually which the dispersal routes between and within the three compartments (humans, farmed animals and the external environment) are ^{6–8}. We also have a reasonably clear picture of what processes result in recruitment of resistance factors from environmental bacteria to human pathogens ⁹. However, much remains to be done in terms of quantifying the relative importance of these processes, which is indispensable knowledge for effective strategies to minimize the spread and development of antibiotic resistance ¹⁰.

It has been proposed that the environment could have three main roles in resistance development ⁹. Firstly, it enables the transfer of resistance genes between environmental, human and animal associated bacteria. Secondly, it constitutes a reservoir or intermediate habitat for resistant bacteria and resistance genes. Finally, it can play a crucial role in the evolution of novel resistance factors, as it provides an arena for selection of resistance combined with an enormous source of genetic diversity

from which bacteria can recruit resistance genes ^{8,9}. Newly emerged resistance genes from the environment can subsequently be disseminated into the human and farmed animals compartments, where they have the potential to cause severe health threats. While this conceptual role of the environment is well described, its actual importance in resistance development is so far unknown, largely due to a lack of quantitative data ^{10,11}. Particularly, the probability of emergence of novel resistance determinants in the environment and their transfer rates to human pathogens are completely unknown this far ¹¹. This means that it becomes very hard to rank risk scenarios and prioritize between possible interventions targeting the emergence of novel resistance factors. The primary aim of this study is to identify the steps within the chain of processes which are limiting the appearance of antibiotic resistance determinants in human pathogens. To address this, we set up a model framework and quantified the dependencies of variables given estimates for upper and lower bounds for the process rates available in literature. Furthermore, we assess how the uncertainty in the available data affects certain parameters in the model, thereby highlighting where the most urgent needs for quantitative data exist.

Methods

Implementation of the model

We consider two different scenarios in our model. In the first scenario, ARGs that later end up in human pathogens already pre-existed in bacteria in some setting at the start of the antibiotic era (around 70 years ago), and only needed to be, in some way, transferred to human pathogens (Table 1). In the second model scenario, we assume that ARGs did not have a resistance function at the start of the modelled time period but needed to first emerge as resistance factors. We will refer to the latter scenario as the “emergence model” and the first scenario as the “pre-existing model”.

Fig.1. Overview of the model framework. Processes are framed in bold. Arrows display the influence of different parameters on the processes. Coloured arrows represent the pathways included in the model (E1 to E6).

The endpoint of the proposed models is a mobilized ARG observed in a human pathogen. Several different chains of events may lead from the first appearance of an ARG to its occurrence in a pathogenic strain (Fig.1). These combinations of events are summarized in the model as six different pathways. First, an ARG may have existed directly on a mobile genetic element in a human pathogen

living in a human being (E1). Secondly, an ARG could have already existed on the chromosome of a human commensal bacteria. After mobilization, the ARG could then be horizontally transferred to a human pathogen (E2). Thirdly, an ARG could originate from a mobile genetic element in a commensal species belonging to the human microbiome. In this scenario, it would then be directly transferrable into a human pathogen via HGT (E3). These first three processes all target pre-existence of ARGs in bacteria directly associated with humans. However, it is also plausible that ARGs may have originated in bacteria living in the environment and subsequently been disseminated to humans. The simplest case would be an ARG that originated from a human pathogen present in the environment. It could then be disseminated between the compartments before ending up in humans (E4). Alternatively, a chromosomally encoded ARG could have existed among environmental bacteria that are not pathogenic. Subsequently, that ARG needs to be mobilized onto a plasmid or as a transposon, spread via horizontal gene transfer (HGT, referred to with the letter 'H' in the model) and be transferred to human pathogens. Humans might pick up the bacterium harbouring the ARG, for instance, with food or while swimming (E5). An ARG could also have pre-existed on a mobile genetic element in an environmental bacterium. Via transformation, transduction or conjugation (H), it could then get transferred to a human pathogen, which can subsequently be taken up by humans (E6).

Time plays an important role in all processes relating to the spread and maintenance of ARGs. For mobilization (M), horizontal gene transfer (H) and the dissemination processes (D), it is important that the microbe carrying the gene survives. Many human pathogenic species are known to have a rather short lifetime outside of their host ^{12,13}. Furthermore, the fitness effect of ARG carriage on population size over time is cumulative, affecting the population expansion rate (S) exponentially, in contrast to the other model parameters. The dissemination efficiency is additionally influenced by the distance to humans in terms of space and time. Human exposure is a critical factor for the dissemination as well as the appearance of the ARG in pathogens in the human microbiome. Humans are exposed to bacteria in many different ways through which uptake via food seems to be the most important one in terms of bacterial exposure per day ¹⁴.

Each of the six pathways shown in Fig. 1 were expressed as equations (Table 1) with parameters defined in Table 2. The interpretation of each equation is the number of mobilized ARGs in pathogens at time t contributed by the corresponding pathway. Both models (the main “pre-existing” model and the “emergence” model) were implemented in R as follows, only differing in the meaning and range of the parameter E (Table 2). Each parameter was randomly selected from its log10 transformed range (see Table 2) under a uniform distribution. For each set of randomly selected parameters, the result of the equations for the different considered processes were calculated (Table 1). Next, the total appearance was calculated based on these sub-equations according to this formula:

$$\text{App}(t) = E1(t) + E2(t) + E3(t) + E4(t) + E5(t) + E6(t)$$

where E1 to E6 each represent the contribution of ARGs by a particular process, expressed as total contributed ARGs over the entire model timeframe, and t is the time in days, resulting in the total appearance (App) representing the total observed ARGs at time t. The true total appearance was approximated from the ResFinder database ¹⁵, which depending on how gene variants are counted held on the order of 700 to 2200 unique ARGs in September 2019. By assuming that these genes would all have appeared in the past 70 years since antibiotics were widely introduced, that gives us an expected appearance rate of around 9 to 29 ARGs per year, or 0.025 to 0.079 per day (Table 2). If the modelled total appearance fell within the expected interval, the model parameters were saved. This procedure was iterated until 10,000 valid sets of model parameters had been obtained for each modelled time point and set of parameter ranges.

Table 1. Considered pathways and the corresponding equations in the two models

Description of pathway	Equation (main model)	Equation (emergence model)
Appearance directly on a mobile genetic element in a human pathogen	$E1(t) = 10^{30} * Pph * Pm * E * S^t$	$E1(t) = 10^{30} * Pph * Pm * E * t * S^{t/2}$
Appearance in non-pathogenic human bacteria, mobilized and transferred to human pathogens	$E2(t) = 10^{30} * Ph * E * MH * t * S^t$	$E2(t) = 10^{30} * Ph * E * t * MH * t/2 * S^{t/2}$
Appearance on a mobile genetic element in non-pathogenic human-associated bacteria	$E3(t) = 10^{30} * Ph * Pm * E * H * t * S^t$	$E3(t) = 10^{30} * Ph * Pm * E * t * H * t/2 * S^{t/2}$
Appearance in pathogens in environment and disseminated to humans	$E4(t) = 10^{30} * Pp * Pm * E * S^t * D * t$	$E4(t) = 10^{30} * Pp * Pm * E * t * S^{t/2} * D * t/2$
Appearance in environmental bacteria, mobilized, transferred to pathogens and disseminated to humans	$E5(t) = 10^{30} * E * MH * t * S^t * D * t/2$	$E5(t) = 10^{30} * E * t * MH * t/2 * S^{t/2} * D * t/4$
Appearance on a mobile genetic element in environmental bacteria, transferred to pathogens and disseminated to humans	$E6(t) = 10^{30} * Pm * E * H * t * S^t * D * t/2$	$E6(t) = 10^{30} * Pm * E * t * H * t/2 * S^{t/2} * D * t/4$

For process and parameter abbreviations, see Table 2. t is for time in days. 10^{30} represents the approximate total number of bacteria on earth.

Table 2. Estimates for the respective process rates and parameters used in the model

Abbreviation	Meaning	Unit	Literature values			Model boundaries	
			Lower bound	Upper bound	Median	Lower bound	Upper bound
E	Proportion of bacteria in which a given ARG exists at the model start or Emergence rate of a novel ARG (in emergence model)	- Events/cell/day	Unknown Unknown			10^{-30} 10^{-40}	1 1
App	First detected appearance of a novel ARG in a human pathogen in a human	Events/day	0.025	0.079	N/A	0.025	0.079
M	Mobilization of a chromosomally encoded ARG onto a plasmid, as a transposon or other conjugative element	Events/cell/day	Difficult to disentangle from horizontal transfer based on current experimental evidence, see MH				
H	Horizontal gene transfer (conjugation)	Events/cell/day	$2.4 * 10^{-10}$	$5.8 * 10^{-2}$	$3.0 * 10^{-3}$	10^{-11}	10^{-1}
D	Dissemination from the environment to humans	Events/cell/day	10^{-14}	10^{-11}	N/A	10^{-15}	10^{-10}
Pph	Fraction of all bacterial cells which are human pathogens and living in/on humans	-	$\sim 10^{-10}$ ($\sim 10^{13}$ bacterial cells per human * $\sim 10^{10}$ humans worldwide * $\sim 10^{-3}$ bacteria living in humans pathogenic / $\sim 10^{30}$ bacterial cells on the world)			10^{-12}	10^{-8}

Ph	Fraction of all bacterial cells that live in/on humans of total bacterial cells on the world	-	$\sim 10^{-7}$ ($\sim 10^{13}$ bacterial cells per human * $\sim 10^{10}$ humans worldwide / $\sim 10^{30}$ bacterial cells on the world)			10^{-8}	10^{-6}
MH	Mobilization of a chromosomally encoded ARG and transfer into another strain	Events/cell/day	$4 * 10^{-10}$	$1.5 * 10^{-4}$	$\sim 5.67 * 10^{-9}$	10^{-15}	10^{-2} ^a
S	Population expansion rate (1 corresponds to no change in population size)	day ⁻¹	Unknown			0.9	1.1
Pm	Probability that a novel ARG emerges on a plasmid	-	0.002 (~ 7% of all bacterial cells carry a conjugative plasmid, most plasmids have ~ 50-100 genes; about 3% of the total bacterial genome)			0.0001	0.01
Pp	Fraction of human pathogenic cells (in all compartments) of bacterial cells on the world	-	10^{-9} , must be larger than Pph, but how much larger is unknown			10^{-11}	10^{-7}

^a MH is also restricted to be < H.

Results

Defining the probability boundaries for antibiotic resistance development

In order to populate the model, measured values for the individual parameters were taken from literature (Supplementary Tables 1 and 2), and the upper and lower boundaries were fed into the model (Table 2). For the rates of horizontal gene transfer (H) most studies were found for the conjugative transfer of genes. Data for both intra- and interspecies transfer of various naturally occurring and artificially constructed plasmids are available. These have been included as a first approximation in the model, bearing in mind that this is laboratory data and that their transferability to the respective ecosystems has been little studied so far. At this time, we have limited knowledge concerning the role of transduction and transformation for the distribution of resistance genes in the respective ecosystems, although recent studies have suggested that they may be important in the transfer of ARGs between bacteria ¹⁶. Due to the lack of quantitative data, these two processes were not considered for the estimation of the parameter H.

Unfortunately, there are no direct measurements for the mobilization parameter (M) that have been obtained independent of HGT. Most studies have the following design in common ¹⁷⁻²¹; the resistance encoded by the donor is not mobile and can only be detected in the recipient once it has been mobilized and transferred. Thus, we do not have a measure of the M parameter, and instead we have used the combined parameter MH (mobilization and horizontal gene transfer) where experimental data can be obtained from these studies.

To get an estimate for the dissemination of ARGs from the environment to humans (D), the number of eaten bacteria per day were used as a proxy ¹⁴. In this study, the authors counted the number of colony forming units of meals for three different diet types. In this context, other transfer routes are also conceivable, e.g., the absorption of bacteria by swallowing water while bathing ^{22,23}. Such dissemination pathways can be of decisive importance when considering individual systems. However, they are not explicitly further considered in this model, which is intentionally kept general, as we aim to accommodate all possible dispersal scenarios in one single parameter.

Finally, some parameters were calculated based on the estimates of the number of humans on earth (7.8 billion; <http://www.worldometers.info>), the number of bacterial cells in/on the human body ($3.8 \times$

10^{13})²⁴ and the total number of bacterial cells on earth (on the order of 10^{30})²⁵. These values were used to define ranges for Pph, Ph, Pp and E (Table 2). We also estimate a typical bacterial genome to contain 1500 to 7500 genes²⁶, that around 7% of bacterial cells carry a conjugable plasmid, and that a typical plasmid carries 50-100 genes²⁷, which was used to constrain the Pm parameter. The S parameter represents the overall fitness impact of carrying an ARG and implicitly also involves the survival rate for a bacterium carrying an ARG, relative to a non-carrier. S is defined as the population expansion rate per day in the model, so when S is equal to one the average ARG would have no impact on the expansion of the population, i.e. the average ARG would overall be fitness neutral.

Fig.2. Valid parameter ranges (A) and process rates (B) for the main model after 70 years of simulated time. The range of S extends from around 0.99 to 1.002.

Many resistance genes are likely to have been present in human-associated bacteria at the beginning of the antibiotic era

It is clear from running the model that for many parameters the permitted range of possible values is huge (Fig. 2A). Even for intuitively ridiculous models such as the assumption that all bacteria originally carried all resistance genes ($E = 1$), there seem to be valid model scenarios. However, it is clear that in those cases, this would have to be compensated by the vast majority of ARGs having a strongly negative effect on fitness, as evidenced by the strong negative correlation between E and S (Fig. 3; Supplementary Fig. 1). As the model results are highly dependent on certain parameters, it is difficult to say what the typical result of the model would be. However, assuming that the dispersal parameter D is in the range between 10^{-12} and 10^{-11} , the most likely origins for the currently circulating resistance genes at the start of the antibiotic era would be on mobile genetic elements (MGEs) in human-associated bacteria (65.6%) and on MGEs in environmental bacteria (28.4%; Table 3). These two origins are substantially more likely than ARGs having been present in pathogens at the start of the antibiotic era (3.07%), coming from the chromosomes of non-pathogenic human-associated bacteria (2.03%) or from the chromosomes of environmental bacteria (0.883%). In almost all runs of the model, the likelihood of ARGs originating from pathogenic bacteria in the environment is less than 0.00001% (Table 3). However, the possible ranges for all the other five processes are highly overlapping (Fig. 2B).

Fig.3. Correlations between the variables in the main model after 70 years of simulated time.

Table 3. Dependency of the modelled processes on the D parameter

Origin	D:10 ⁻¹⁵ to 10 ⁻¹⁴	D:10 ⁻¹⁴ to 10 ⁻¹³	D:10 ⁻¹³ to 10 ⁻¹²	D:10 ⁻¹² to 10 ⁻¹¹	D:10 ⁻¹¹ to 10 ⁻¹⁰
pathogens in the human microbiome	3.31%	3.74%	3.88%	3.07%	0.733%
non-pathogenic human-associated bacteria	2.81%	3.19%	2.63%	2.03%	0.530%
chromosomes of environmental bacteria	0.00116%	0.123%	0.103%	0.883%	2.18%
MGEs in human-associated bacteria	93.8%	92.7%	89.9%	65.6%	18.9%
MGEs in the environment	0.0389%	0.358%	3.50%	28.4%	77.7%
pathogenic bacteria in the environment	<0.00001%	<0.00001%	<0.00001%	<0.00001%	0.00001%

Furthermore, the interdependencies between the process rates and some of the individual parameters are, in many cases, fairly high. For example, the likelihood of an ARG originating from the chromosomes of environmental bacteria or from non-pathogenic human-associated bacteria is strongly correlated to the value of MH – the mobilization rate of genes to MGEs, including subsequent gene transfer (Fig. 3). Similarly, the proportion of ARGs that already existed in pathogens in humans to begin with is, unsurprisingly, dependent on the proportion of all bacteria that are human pathogens (Pph), and the likelihood of originating on an MGE – either in a human-associated bacterium or in the environment – is strongly dependent on the rate of horizontal gene transfer (H). Importantly, though, the number of bacteria carrying a certain ARG from the start of the modelling time (E) is almost entirely dependent on the fitness cost or advantage associated with the gene (S), highlighting that selection for ARGs is by far the most important process in ARG ecology. The parameter that largely controls whether the environment is likely to be a significant contributor to ARGs in pathogens is D (Fig. 3). If the environment played a major role in the emergence of the nowadays known ARGs in human pathogens,

dissemination from the environment to humans (D) as well as the mobilization and transfer rates (MH) of ARGs need to be fairly high. It is also notable that the three parameters that influence the outcomes of the model the most are S, H and MH (and indirectly E, as it is anti-correlated with S; Fig. 3). In particular, changing the values of MH and H shifts the relative importance of the processes and thus the most likely origin of ARGs dramatically (Supplementary Table 1). Again, it should be pointed out that the span of possible outcomes for a given size of MH, H or D is still very large (Supplementary Figs. 2-4).

Stability of the parameters to changes

We also investigated how stable the model was over time (i.e. when the number of days were varied in the model). Again, the strong relationship between the E and S parameters was evident (Fig. 4A and B). Particularly, the value of S becomes highly constrained over long timescales, hinting at that most ARGs that have made it all the way to human pathogens should be close to neutral with respect to impact on host fitness. It is also interesting to note the relative increase in environmental processes over time (Fig. 4C), which also relates to the impact of selection over time and how the environment presents a vastly larger number of bacteria with potentially neutral ARGs to recruit from.

Fig. 4. Dependence of model process rates (A) and key model parameters E (B) and S (C) over the simulated time span from the start of antibiotic use to the emergence of ARGs in human pathogens. Dotted lines in (A) and dashed lines in (B) and (C) represent the range in which 95% of the values fall in the simulations. Lines with dots and dashes in (B) and (C) represent the range in which 50% of the values fall in the simulations.

Building on the assumption of neutral impact on host fitness (or, more precisely, neutral *average* impact on fitness over the run time of the model), we also ran the model with S fixed to 1 (i.e. no effects on fitness for any ARG). This is an interesting scenario as it caps E to the range between 10^{-24} and 10^{-11} (Supplementary Fig. 6), meaning that if fitness cost is neutral, currently observed ARGs in pathogens would have been present in somewhere between 10^6 and 10^{19} bacterial cells at the start of the antibiotic era. However, other than that fixing S to 1 changes how the different processes depend on the E parameter (Supplementary Fig. 7), the model output does not substantially change compared to running the model with a variable S parameter.

We also compared the main model to the emergence model, in which ARGs are assumed not to have been already present in bacteria at the start of the antibiotic era, but instead had to develop their resistance function – emerge as ARGs – within that time frame. However, the results differed only slightly from the results of the pre-existing model (Supplementary Fig. 8). Based on the model results, the conceptual difference between the pre-existing model and emergence model is not sufficient to change the dependencies between the rest of the processes. The full results from the emergence model are available in the supplementary material.

Model predictions

Notably, the model makes some predictions that may be testable within the foreseeable future. The respective results will substantially contribute to determine the crucial processes in antibiotic resistance development and thus help in the identification of effective intervention strategies:

- (1) Fitness costs of ARGs: The model predicts that the majority of ARGs present in pathogens today should have very limited effects on fitness. The model caps the average fitness impact for ARGs currently present in human pathogens between -0.2% and +0.1% per generation, and 30 years into the future the model predicts that the range of possible fitness effects will have narrowed even further (-0.15% to +0.1%; data not shown). By determining the fitness effects of carrying individual ARGs in their current hosts considering inter- as well as intra-species variability, the prediction that most ARGs impose a very minor fitness impact could be experimentally tested, although it would have to be tested across a large number of ARGs.
- (2) The origin of ARGs: The model predicts the most likely location of ARGs 70 years ago would have been in human-associated bacteria (Table 3). By tracking ARGs currently present in human pathogens across large datasets of bacterial genomes it may be possible to trace the evolutionary history of these genes and thereby identify their likely hosts at the beginning of the antibiotic era. Such an attempt was recently made, corroborating the findings based on our model²⁸ and lending support to that most ARGs may not originate in the environment. However, this analysis is complicated by the biased sampling of fully sequenced bacterial genomes, most of which originate from human specimens (<https://gold.jgi.doe.gov/distribution>). Thus, it could be expected that when performed today, such an analysis would only be able to trace origins for ARGs that were present in human-associated bacteria 70 years ago. Notably, the model estimates that such genes would be more likely than not to make up the majority of all ARGs currently present in pathogens. Importantly, the rapid increase in sequencing capacity may make full-scale analysis of ARG origins using genomic data possible in the relatively near future, which would enable further testing of this prediction of the model.
- (3) More data will enable additional specific predictions: Given that the origins of ARGs currently circulating in pathogens can be established, the range of possible predicted values of D narrows dramatically. If it can be shown that the proportion of ARGs in pathogens today that were already present in human-associated bacteria 70 years ago is not even close to 60%, the dispersal parameter D has to be above 10^{-12} (Table 3). Conversely, if M and H could be better determined by experiments, the model would predict the likely origins more precisely (Supplementary Table 1). Just establishing a ball-park range of the mobilization rate (M) would dramatically restrict the possible outcomes of the model, particularly if M turns out to be at the extreme ends of the spectrum. Thus, a more precise determination of any of these parameters would enable several more specific predictions by the model,

Discussion

In this paper, we present a model for ARG appearance in pathogens, based on the limited data currently available on process rates in the literature. The model predicts a vast range of possible outcomes and parameter values; however, it serves its purpose to highlight which processes and parameters that are interdependent and which ones are key to get a better understanding of how to mitigate antibiotic resistance development. A key aspect shaping the model is the strong relation between longevity of ARGs (intrinsically largely determined by their fitness cost) and how widespread resistance genes were across bacteria at the start of the antibiotic era (which likely serves as a proxy for how large the total genetic reservoir to recruit ARGs from still is today²⁹). As the average fitness cost of carrying an ARG increases in the model (resulting in lower values of the S parameter), the number of available ARGs – either in the human microbiome, the environment, or both – needs to get bigger to compensate for that. As the model scenario plays out over a vast amount of time (in microbial terms), the possible range of average fitness costs or gains associated with ARG carriage becomes very small. Thus, the model essentially predicts that most ARGs present in pathogens today should have no or little fitness cost,

which is in accordance with theoretical arguments we have put forth previously^{9,11}. As mentioned earlier, this prediction is, in principle, testable by large-scale experiments on the impacts of ARG carriage on bacterial fitness. Notably, low average fitness costs for ARGs are associated with a greater likelihood that they have been present in human pathogens all along. Overall, it is clear that the fitness cost of ARGs is a key property for their ability to make their way to and persist in human pathogens. The persistence is likely to be facilitated by exposure to low levels of antibiotics, but also by mutational processes that lead to domestication of ARGs and MGEs, as well as other forms of fitness cost reductions, some of which may yet be unknown. In this context, a variety of mechanisms to reduce the fitness cost of ARGs are known, but quantitative information on their occurrence and effect is still largely lacking (Table 4). More research is needed to quantitatively determine the impact of domestication and reduction of fitness costs on selection of ARGs, as well as what additional mechanisms that are available to bacteria for reducing fitness costs of ARGs. Furthermore, minimal selective concentrations must be determined for a much larger set of antibiotics, environmental conditions and bacterial species, alone as well as in microbial communities, to provide actual scientific data in a field that currently relies largely on predictions of antibiotic effects on microbes^{30,31}.

Table 4. Identified knowledge gaps and research needs that must be addressed to build quantitative risk assessment models and better mitigate ARG recruitment

Knowledge gap	Research need	Utility
Settings that select for antibiotic resistance	Minimal selective concentrations (MSCs) for more antibiotics, environmental conditions and bacterial species, alone, in co-culture and in complete communities	Better sewage and wastewater treatment strategies; emission limits and environmental quality standards for antibiotics
Rate of bacterial dispersal between environmental compartments (and to humans)	Quantitative observations of bacterial dispersal between environments; tracing spread of ARGs between environments over time; better determination of human exposure to environmental bacteria	Ability to discern if the environment has a significant role in the emergence of novel ARGs; improved quantitative risk assessment models for environmental antibiotic resistance; ability to limit spread of ARGs and resistant bacteria from the environment to humans; improved wastewater treatment strategies
Quantitative information on the occurrence and effect of mechanisms for fitness cost reduction and domestication of ARGs	Quantitative understanding of genomic mechanisms and genes responsible for fitness cost reduction of ARGs	Improved quantitative risk assessment models for environmental antibiotic resistance
Rate of horizontal gene transfer between bacteria	Measurement of transfer rates (especially for transformation and transduction) between a larger diversity of bacterial species and in a greater number of settings	Improved quantitative risk assessment models for environmental antibiotic resistance; ability to curb transfer of ARGs between bacteria or reduce the number of settings where this can take place
Rate of mobilization of genes from bacterial chromosomes to MGEs	Development of proper experimental setups to measure mobilization; measurement of mobilization rates in such assays	Improved quantitative risk assessment models for environmental antibiotic resistance; better understanding of whether the environment acts as a source of ARGs to human pathogens, and thereby better prioritization of interventions

Along with the importance of fitness costs and selective effects, the model also highlights the fundamentally important impact of the dispersal parameter (D) on ARG occurrence in pathogens. This parameter essentially determines if the environment would at all plausibly function as a reservoir for ARGs to be recruited to pathogens, as has been argued by many authors^{4,32–34}. Notably, there are

more scenarios in the model where the vast majority of ARGs originated in human-associated bacteria than there are scenarios where the environment played a significant role, although both situations are possible. Interestingly, a recent study using genomic information to infer the recent origins of ARGs assigned more than 90% to bacteria associated with humans and/or domestic animals²⁸, which is consistent with the most common results of our model. While there are some limitations to the currently available genomic data, this implies that the environment would only play a minor role in the development of new ARGs and, consequently, that the D parameter is likely to be smaller than 10^{-12} .

Importantly, the data on dispersal of genes from environmental bacteria to the human population is very limited, particularly for the type of generalized case that we outline in this paper. Due to the lack of data, we have chosen a fairly naive definition of dispersal for our model, reflecting the type of data that is available. The D parameter in the model represents the proportion of environmental bacteria that would be expected to be exposed to a human somewhere in the world during a day. This was estimated from the number of cultivable bacteria contained in different types of food as reported by Lang et al.¹⁴. However, depending on the diet investigated, the approximate numbers of bacteria taken in from food were in the range of 10^6 to 10^9 bacteria per person and day, which is a fairly wide range. In addition, this does not account for bacterial exposure through other routes, such as recreational swimming^{22,23}, wounds³⁵ and through direct contact with animals^{33,36,37}. Thus, the definition of D in our model does not reflect that the origins of bacteria that the average human is exposed to are not equally distributed across environments. To some extent, we expect this to be compensated by bacterial global dispersal over time, but this still reflects a weakness in our model. That said, although it is feasible that all these activities contribute significant numbers of bacteria, it is unlikely that they would change D by several orders of magnitude. However, for this reason it may be fair to assume that D is in the higher end of the spectrum, i.e. larger than 10^{-12} rather than close to the lower boundary. If D would be larger than 10^{-11} , environmental bacteria may have contributed around 80% of all currently clinically relevant resistance genes (Table 3). This, however, runs contrary to recent genomic evidence²⁸, although the latter is based on a likely biased sample of genomic data. The impact of the D parameter highlights that determining the actual dispersal rate of ARGs from the environment is crucial to understand the role of the environment in antibiotic resistance development. Comprehensive data on ARG dispersal would not only aid in risk assessment of emergence of novel resistance factors; it would also aid in understanding and mitigating the spread of resistant bacteria (and their ARGs) through the environment^{6,8,10,38}.

The model also reveals that HGT and mobilization of chromosomal ARGs to MGEs are potentially important processes, although likely less to be bottlenecks in resistance development than selection and dispersal. While we have a hunch about what the HGT rates may be (ranging from 2.4×10^{-10} to 5.8×10^{-2} events per cell per day in our literature review; Supplementary Table S2), this range is still not sufficiently narrow to really restrict the model outcomes in a meaningful way. That said, at high rates of HGT, most ARGs are, according to the model, likely to have originated on MGEs, unless the mobilization rate is also high. Unfortunately, the latter parameter is almost completely unknown – the little available data we could find suggest that the rate of mobilization followed by HGT is in the range between 4×10^{-10} and 1.5×10^{-4} . Since these estimates are all based on mobilization followed by HGT, they do not represent the mobilization rate *per se*, and thus the actual rate may turn out to be outside of the mentioned range. In addition, these rates have been measured using very specific markers that are known to be transferable to plasmids, which may not make them representative for the mobilization rate of ARGs in general. Clearly, there is a need to better define the typical rates of HGT between bacteria. Several assays already exist to perform such experiments^{39–42}, but the amount of data available is still insufficient as it is only available for a very limited number of conditions, ARGs and mating pairs. Furthermore, most research on HGT has dealt with conjugation, while transformation and transduction has been largely neglected, particularly from a quantitative point of view. Still, the two latter processes have been shown to be important drivers of the horizontal transfer of resistance genes^{16,43–45}, although their contribution relative to conjugation is unclear. Importantly, this is largely a data

generation problem. In contrast, the lack of knowledge of the rate of mobilization of chromosomal genes to MGEs is not only related to insufficient data, but also to a lack of appropriate assays to actually measure mobilization of ARGs rather than mobilization linked to, e.g., horizontal gene transfer. This highlights an urgent need for assays to directly detect mobilization of genes from the chromosomes to MGEs.

It is important to note that we have intentionally kept the model and its parameters general, in order to provide an overview of the respective importance of the individual processes that influence the appearance of new resistance genes in human pathogens. This is likely also one of the reasons why the model predictions are in many cases associated with very wide ranges. However, the model still emphasizes important and interrelated processes that need more research attention in order to build more quantitatively accurate models. Furthermore, by using specific rates for particular ecosystems as input for the model, we believe that our model could be used to assess the probabilities for the contribution of specific pathways for resistance development and dissemination. Importantly, this model constitutes an initial effort to create a conceptual framework for ARG emergence in pathogens^{9,11} that is quantitative, and to highlight where more data is needed in order to make more precise predictions. As such, we hope that this modelling effort can spur more research; both into determining the model parameter ranges more precisely, but also into more sophisticated modelling approaches that may better reveal the details on ARG emergence.

Another limitation to the model is that the variable associated with the fitness cost of ARGs (S) represents a variety of different processes and parameters related to bacterial survival and proliferation. As we did not find an unambiguous way to tease these components apart, they are all captured by this single parameter, which of course results in an oversimplification of the real world. For these reasons, the role of S has to be interpreted with some caution. First of all, S represents the *average* fitness impact of the *average* ARG over the entire run time of the model (which in most of the data we have presented in this paper is 70 years). This means that S encapsulates scenarios where an ARG may some of the time have a positive fitness impact (such as during antibiotic selection) and at other times a slightly negative effect on fitness (e.g., in the absence of antibiotic selection). The final value of S therefore represents the expected overall fitness impact of ARG carriage, not the specific fitness of any single ARG currently encountered in pathogens. Furthermore, S also implicitly includes a constraint on the survival rate for a bacterium carrying an ARG. The way S is defined in the model, it is related to the population expansion rate per generation. In other words, if S is equal to one, the average ARG is expected to have no impact on bacterial growth and, thus, the population expansion (i.e. it is fitness neutral). However, given the relatively quick doubling rate of bacteria, a value only slightly deviating from one will rapidly result in the gene propagating through bacterial populations or quickly disappearing, which explains the tendency of S to approach one the longer the simulated time of the model.

Despite the large uncertainties associated with the model results, some useful intervention priorities can still be derived from this work (Table 4). Firstly, due to the strong influence of selection and ARG fitness costs on the model outcomes, it is clear that there is a need for a deeper understanding of the settings that drive antibiotic resistance selection. Identifying threshold concentrations of antibiotics (and other selective agents, such as biocides^{27,46}) that select or co-select for antibiotic resistance is crucial for determining which emission limits that should be imposed on effluents from pharmaceutical production^{47,48} as well as from regular wastewater treatment^{49–51}. Currently, regulatory efforts are mostly based on predicted no-effect concentrations^{30,31}, and while this is a decent interim solution, solid scientific data on actual minimal selective concentrations in relevant settings are urgently needed¹⁰.

Filling the knowledge gaps in terms of human exposure to environmental bacteria carrying ARGs also has important implications for prioritization of interventions to prevent the emergence of antibiotic resistance in pathogens. The data we present in this paper suggest that a low dispersal rate renders

the environment implausible as a source for the vast majority of ARGs in pathogens. If this is true, that would imply that in the short term, interventions targeting human use of antibiotics should be prioritized over, e.g., environmental interventions, as the latter might be ineffective in preventing resistance emergence in pathogens. That said, the environment may still function as a dispersal route for already resistant pathogens, and thus environmental mitigations should be selectively targeted towards minimizing spread of resistant bacteria from the human population to the environment.

The model furthermore implies that a low mobilization rate would cause the spread of ARGs from resistant bacteria to susceptible pathogens to be the most important vehicle for recruitment of novel ARGs. This may also be the case if mobilization rates are high, but then to a lesser extent. These observations highlight the importance of improving our determination of the typical range of HGT rates of ARGs between bacteria. The impact of environmental conditions and stressors on the transfer rate is also crucial to determine, as that knowledge would potentially allow interventions to reduce the number of settings where transfer between bacteria can take place, and thus delay resistance emergence in pathogens.

Finally, if we can gain a better understanding of how resistant bacteria and ARGs disseminate via the environment between compartments and, particularly, from the environment to humans, it would be possible to start designing mitigation strategies aiming to reduce the spread of resistant bacteria through the environmental route. Such interventions could be made either on the level of releases into the environment or attempting to create barriers between humans and particular risk settings⁸. Which of those that would be most beneficial is to a large extent determined by the rate of dispersal of resistant bacteria from the environments to humans. Sadly, our current knowledge of the environmental dissemination routes is poor, mostly anecdotal and virtually without quantitative measurements, which is also reflected in the simplistic definition of the dispersal parameter in our model.

Conclusions and outlook

In this paper, we present a model to describe the emergence of ARGs in human pathogens, and populate it using the limited literature data that exist. The model indicates that fitness costs of ARG carriage is one of the key factors determining the fate of ARGs in both the environment and in human pathogens. It also points to the importance of, particularly, the degree of dispersal of resistant bacteria from the environment to humans, but also the rates of mobilization and horizontal transfer of ARGs. All of these processes are poorly described at present; particularly the possible range of mobilization rates is practically unknown and estimates of dispersal of bacteria from the environment to humans are uncertain. Furthermore, depending on how large these rates are, the model predicts very different probabilities for the possible origins of the ARGs we currently see in pathogens, indicating yet again the importance of determining these values with better precision. However, despite a large degree of uncertainty of specific model results, the model clearly highlights where the most urgent needs for quantitative data exist. The knowledge that could be gained from more precise quantitative process data would be extremely helpful in deciding which processes that should be targeted for interventions of environmental antibiotic resistance development and transmission. By measuring process rates for mobilization, transfer and dissemination, as well as the fitness costs associated with ARG carriage, the model used here could be refined to make more precise predictions on the importance of certain parameters and environments. This will allow for better monitoring of antibiotic resistance and interventions towards preventing it in the future.

Acknowledgements

JBP acknowledges funding from the Centre for Antibiotic Resistance Research at the University of Gothenburg, the Swedish Research Council for Environment, Agricultural Sciences and Spatial Planning (FORMAS; grant 2016-00768), the Swedish Research Council (VR; grant 2019-00299) under the frame of JPI AMR (EMBARK; JPIAMR2019-109), the Sahlgrenska Academy at the University of

Gothenburg, and the Swedish Cancer and Allergy fund (Cancer- och Allergifonden). SH acknowledges funding from the German Research Foundation (HE 8047/3-1)

References

- (1) Review on Antimicrobial Resistance. *Tackling Drug-Resistant Infections Globally: Final Report and Recommendations*; 2016.
- (2) European Commission. *A European One Health Action Plan against Antimicrobial Resistance (AMR)*; European Commission, 2017.
- (3) Ashbolt, N. J.; Amézquita, A.; Backhaus, T.; Borriello, P.; Brandt, K. K.; Collignon, P.; Coors, A.; Finley, R.; Gaze, W. H.; Heberer, T.; Lawrence, J. R.; Larsson, D. G. J.; McEwen, S. A.; Ryan, J. J.; Schönfeld, J.; Silley, P.; Snape, J. R.; Van den Eede, C.; Topp, E. Human Health Risk Assessment (HHRA) for Environmental Development and Transfer of Antibiotic Resistance. *Environmental health perspectives* **2013**, *121* (9), 993–1001. <https://doi.org/10.1289/ehp.1206316>.
- (4) Finley, R. L.; Collignon, P.; Larsson, D. G. J.; McEwen, S. A.; Li, X.-Z.; Gaze, W. H.; Reid-Smith, R.; Timinouni, M.; Graham, D. W.; Topp, E. The Scourge of Antibiotic Resistance: The Important Role of the Environment. *Clinical Infectious Diseases* **2013**, *57* (5), 704–710. <https://doi.org/10.1093/cid/cit355>.
- (5) Gaze, W. H.; Krone, S. M.; Larsson, D. G. J.; Li, X.-Z.; Robinson, J. A.; Simonet, P.; Smalla, K.; Timinouni, M.; Topp, E.; Wellington, E. M.; Wright, G. D.; Zhu, Y.-G. Influence of Humans on Evolution and Mobilization of Environmental Antibiotic Resistome. *Emerging infectious diseases* **2013**, *19* (7). <https://doi.org/10.3201/eid1907.120871>.
- (6) Pruden, A.; Larsson, D. G. J.; Amézquita, A.; Collignon, P.; Brandt, K. K.; Graham, D. W.; Lazorchak, J. M.; Suzuki, S.; Silley, P.; Snape, J. R.; Topp, E.; Zhang, T.; Zhu, Y.-G. Management Options for Reducing the Release of Antibiotics and Antibiotic Resistance Genes to the Environment. *Environmental health perspectives* **2013**, *121* (8), 878–885. <https://doi.org/10.1289/ehp.1206446>.
- (7) Berendonk, T. U.; Manaia, C. M.; Merlin, C.; Fatta-Kassinos, D.; Cytryn, E.; Walsh, F.; Bürgmann, H.; Sørum, H.; Norström, M.; Pons, M.-N.; Kreuzinger, N.; Huovinen, P.; Stefani, S.; Schwartz, T.; Kisand, V.; Baquero, F.; Martínez, J. L. Tackling Antibiotic Resistance: The Environmental Framework. *Nature reviews Microbiology* **2015**, *13* (5), 310–317. <https://doi.org/10.1038/nrmicro3439>.
- (8) Bengtsson-Palme, J.; Hess, S. Strategies to Reduce or Eliminate Resistant Pathogens in the Environment. In *Antimicrobial Drug Resistance*; Wiley, 2019.
- (9) Bengtsson-Palme, J.; Kristiansson, E.; Larsson, D. G. J. Environmental Factors Influencing the Development and Spread of Antibiotic Resistance. *FEMS Microbiology Reviews* **2018**, *42* (1), 25. <https://doi.org/10.1093/femsre/fux053>.
- (10) Larsson, D. G. J.; Andremont, A.; Bengtsson-Palme, J.; Brandt, K. K.; de Roda Husman, A. M.; Fagerstedt, P.; Fick, J.; Flach, C.-F.; Gaze, W. H.; Kuroda, M.; Kvint, K.; Laxminarayan, R.; Manaia, C. M.; Nielsen, K. M.; Plant, L.; Ploy, M.-C.; Segovia, C.; Simonet, P.; Smalla, K.; Snape, J.; Topp, E.; van Hengel, A. J.; Verner-Jeffreys, D. W.; Virta, M. P. J.; Wellington, E. M.; Wernersson, A.-S. Critical Knowledge Gaps and Research Needs Related to the Environmental Dimensions of Antibiotic Resistance. *Environment International* **2018**, *117*, 132–138. <https://doi.org/10.1016/j.envint.2018.04.041>.
- (11) Bengtsson-Palme, J. Assessment and Management of Risks Associated With Antibiotic Resistance in the Environment. In *Management of Emerging Public Health Issues and Risks*; Elsevier, 2019; pp 243–263. <https://doi.org/10.1016/B978-0-12-813290-6.00010-X>.
- (12) Kusumaningrum, H. D.; Riboldi, G.; Hazeleger, W. C.; Beumer, R. R. Survival of Foodborne Pathogens on Stainless Steel Surfaces and Cross-Contamination to Foods. *Int J Food Microbiol* **2003**, *85* (3), 227–236. [https://doi.org/10.1016/s0168-1605\(02\)00540-8](https://doi.org/10.1016/s0168-1605(02)00540-8).
- (13) Bitton, G.; Farrah, S. R.; Ruskin, R. H.; Butner, J.; Chou, Y. J. Survival of Pathogenic and Indicator Organisms in Ground Water. *Groundwater* **1983**, *21* (4), 405–410. <https://doi.org/10.1111/j.1745-6584.1983.tb00741.x>.
- (14) Lang, J. M.; Eisen, J. A.; Zivkovic, A. M. The Microbes We Eat: Abundance and Taxonomy of Microbes Consumed in a Day's Worth of Meals for Three Diet Types. *PeerJ* **2014**, *2*, e659. <https://doi.org/10.7717/peerj.659>.
- (15) Bortolaia, V.; Kaas, R. S.; Ruppe, E.; Roberts, M. C.; Schwarz, S.; Cattoir, V.; Philippon, A.; Allesoe, R. L.; Rebelo, A. R.; Florensa, A. F.; Fagelhauer, L.; Chakraborty, T.; Neumann, B.; Werner, G.; Bender, J. K.; Stingl, K.; Nguyen, M.; Coppens, J.; Xavier, B. B.; Malhotra-Kumar, S.; Westh, H.; Pinholt, M.; Anjum, M. F.; Duggett, N. A.; Kempf, I.; Nykäsenoja, S.; Olkkola, S.; Wiczorek, K.; Amaro, A.; Clemente, L.; Mossong, J.; Losch, S.; Ragimbeau, C.; Lund, O.; Aarestrup, F. M. ResFinder 4.0 for Predictions of Phenotypes from Genotypes. *Journal of Antimicrobial Chemotherapy* **2020**, *75* (12), 3491–3500. <https://doi.org/10.1093/jac/dkaa345>.
- (16) Crippen, C. S.; Jr., M. J. R.; Sanchez, S.; Szymanski, C. M. Multidrug Resistant Acinetobacter Isolates Release Resistance Determinants Through Contact-Dependent Killing and Bacteriophage Lysis. *Frontiers in Microbiology* **2020**, *11*. <https://doi.org/10.3389/fmicb.2020.01918>.
- (17) Franke, A. E.; Clewell, D. B. Evidence for a Chromosome-Borne Resistance Transposon (Tn916) in *Streptococcus Faecalis* That Is Capable of “Conjugal” Transfer in the Absence of a Conjugative Plasmid. *Journal of Bacteriology* **1981**, *145* (1), 494–502. <https://doi.org/10.1128/JB.145.1.494-502.1981>.

- (18) Scott, J. R.; Kirchman, P. A.; Caparon, M. G. An Intermediate in Transposition of the Conjugative Transposon Tn916. *Proceedings of the National Academy of Sciences of the United States of America* **1988**, 85 (13), 4809–4813.
- (19) Torres, O. R.; Korman, R. Z.; Zahler, S. A.; Dunny, G. M. The Conjugative Transposon Tn925: Enhancement of Conjugal Transfer by Tetracycline in *Enterococcus Faecalis* and Mobilization of Chromosomal Genes in *Bacillus Subtilis* and *E. Faecalis*. *Molecular & general genetics : MGG* **1991**, 225 (3), 395–400.
- (20) Poyart, C.; Celli, J.; Trieu-Cuot, P. Conjugative Transposition of Tn916-Related Elements from *Enterococcus Faecalis* to *Escherichia Coli* and *Pseudomonas Fluorescens*. *Antimicrobial Agents and Chemotherapy* **1995**, 39 (2), 500–506. <https://doi.org/10.1128/AAC.39.2.500>.
- (21) Manson, J. M.; Hancock, L. E.; Gilmore, M. S. Mechanism of Chromosomal Transfer of *Enterococcus Faecalis* Pathogenicity Island, Capsule, Antimicrobial Resistance, and Other Traits. *Proceedings of the National Academy of Sciences of the United States of America* **2010**, 107 (27), 12269–12274. <https://doi.org/10.1073/pnas.1000139107>.
- (22) Leonard, A. F. C.; Zhang, L.; Balfour, A. J.; Garside, R.; Gaze, W. H. Human Recreational Exposure to Antibiotic Resistant Bacteria in Coastal Bathing Waters. *Environment International* **2015**, 82, 92–100. <https://doi.org/10.1016/j.envint.2015.02.013>.
- (23) Leonard, A. F. C.; Zhang, L.; Balfour, A. J.; Garside, R.; Hawkey, P. M.; Murray, A. K.; Ukoumunne, O. C.; Gaze, W. H. Exposure to and Colonisation by Antibiotic-Resistant *E. Coli* in UK Coastal Water Users: Environmental Surveillance, Exposure Assessment, and Epidemiological Study (Beach Bum Survey). *Environment International* **2018**, 114, 326–333. <https://doi.org/10.1016/j.envint.2017.11.003>.
- (24) Sender, R.; Fuchs, S.; Milo, R. Revised Estimates for the Number of Human and Bacteria Cells in the Body. **2016**, 14 (8), e1002533. <https://doi.org/10.1371/journal.pbio.1002533>.
- (25) Kallmeyer, J.; Pockalny, R.; Adhikari, R. R.; Smith, D. C.; D'Hondt, S. Global Distribution of Microbial Abundance and Biomass in Subseafloor Sediment. *Proceedings of the National Academy of Sciences of the United States of America* **2012**, 109 (40), 16213–16216. <https://doi.org/10.1073/pnas.1203849109>.
- (26) Gregory, T. R. Synergy between Sequence and Size in Large-Scale Genomics. *Nat Rev Genet* **2005**, 6 (9), 699–708. <https://doi.org/10.1038/nrg1674>.
- (27) Pal, C.; Bengtsson-Palme, J.; Kristiansson, E.; Larsson, D. G. J. Co-Occurrence of Resistance Genes to Antibiotics, Biocides and Metals Reveals Novel Insights into Their Co-Selection Potential. *BMC Genomics* **2015**, 16 (1), 964. <https://doi.org/10.1186/s12864-015-2153-5>.
- (28) Ebmeyer, S.; Kristiansson, E.; Larsson, D. G. J. A Framework for Identifying the Recent Origins of Mobile Antibiotic Resistance Genes. *Communications Biology* **2021**, 4 (1). <https://doi.org/10.1038/s42003-020-01545-5>.
- (29) Bengtsson-Palme, J.; Larsson, D. G. J. Antibiotic Resistance Genes in the Environment: Prioritizing Risks. *Nature reviews Microbiology* **2015**, 13 (6), 396. <https://doi.org/10.1038/nrmicro3399-c1>.
- (30) Bengtsson-Palme, J.; Larsson, D. G. J. Concentrations of Antibiotics Predicted to Select for Resistant Bacteria: Proposed Limits for Environmental Regulation. *Environment International* **2016**, 86, 140–149. <https://doi.org/10.1016/j.envint.2015.10.015>.
- (31) Tell, J.; Caldwell, D. J.; Häner, A.; Hellstern, J.; Hoeger, B.; Journal, R.; Mastrocco, F.; Ryan, J. J.; Snape, J.; Straub, J. O.; Vestel, J. Science-Based Targets for Antibiotics in Receiving Waters from Pharmaceutical Manufacturing Operations. *Integrated Environmental Assessment and Management* **2019**, 15 (3), 312–319. <https://doi.org/10.1002/ieam.4141>.
- (32) Martinez, J. L. Antibiotics and Antibiotic Resistance Genes in Natural Environments. *Science (New York, NY)* **2008**, 321 (5887), 365–367. <https://doi.org/10.1126/science.1159483>.
- (33) Allen, H. K.; Donato, J.; Wang, H. H.; Cloud-Hansen, K. A.; Davies, J.; Handelsman, J. Call of the Wild: Antibiotic Resistance Genes in Natural Environments. *Nature reviews Microbiology* **2010**, 8 (4), 251–259. <https://doi.org/10.1038/nrmicro2312>.
- (34) D'Costa, V. M.; King, C. E.; Kalan, L.; Morar, M.; Sung, W. W. L.; Schwarz, C.; Froese, D.; Zazula, G.; Calmels, F.; Debruyne, R.; Golding, G. B.; Poinar, H. N.; Wright, G. D. Antibiotic Resistance Is Ancient. *Nature* **2011**, 477 (7365), 457–461. <https://doi.org/10.1038/nature10388>.
- (35) Baumgardner, D. J. Soil-Related Bacterial and Fungal Infections. *J Am Board Fam Med* **2012**, 25 (5), 734–744. <https://doi.org/10.3122/jabfm.2012.05.110226>.
- (36) Argudín, M. A.; Deplano, A.; Meghraoui, A.; Dodémont, M.; Heinrichs, A.; Denis, O.; Nonhoff, C.; Roisin, S. Bacteria from Animals as a Pool of Antimicrobial Resistance Genes. *Antibiotics (Basel, Switzerland)* **2017**, 6 (2). <https://doi.org/10.3390/antibiotics6020012>.
- (37) Iramiot, J. S.; Kajumbula, H.; Bazira, J.; Kansime, C.; Asiimwe, B. B. Antimicrobial Resistance at the Human-Animal Interface in the Pastoralist Communities of Kasese District, South Western Uganda. *Sci Rep* **2020**, 10 (1), 14737. <https://doi.org/10.1038/s41598-020-70517-w>.
- (38) Pruden, A.; Alcalde, R. E.; Alvarez, P. J. J.; Ashbolt, N.; Bischel, H.; Capiro, N. L.; Crossette, E.; Frigon, D.; Grimes, K.; Haas, C. N.; Ikuma, K.; Kappell, A.; LaPara, T.; Kimbell, L.; Li, M.; Li, X.; McNamara, P.; Seo, Y.; Sobsey, M. D.; Sozzi, E.; Navab-Daneshmand, T.; Raskin, L.; Riquelme, M. V.; Vikesland, P.; Wigginton, K.; Zhou, Z. An Environmental Science and Engineering Framework for Combating Antimicrobial Resistance. *Environmental Engineering Science* **2018**, ees.2017.0520. <https://doi.org/10.1089/ees.2017.0520>.

- (39) Sun, D.; Zhang, Y.; Mei, Y.; Jiang, H.; Xie, Z.; Liu, H.; Chen, X.; Shen, P. Escherichia Coli Is Naturally Transformable in a Novel Transformation System. *FEMS Microbiol Lett* **2006**, *265* (2), 249–255. <https://doi.org/10.1111/j.1574-6968.2006.00503.x>.
- (40) Etchuuya, R.; Ito, M.; Kitano, S.; Shigi, F.; Sobue, R.; Maeda, S. Cell-to-Cell Transformation in Escherichia Coli: A Novel Type of Natural Transformation Involving Cell-Derived DNA and a Putative Promoting Pheromone. *PLoS One* **2011**, *6* (1), e16355. <https://doi.org/10.1371/journal.pone.0016355>.
- (41) Jutkina, J.; Rutgersson, C.; Flach, C.-F.; Larsson, D. G. J. An Assay for Determining Minimal Concentrations of Antibiotics That Drive Horizontal Transfer of Resistance. *The Science of the total environment* **2016**, *548–549*, 131–138. <https://doi.org/10.1016/j.scitotenv.2016.01.044>.
- (42) Kneis, D.; Berendonk, T. U.; Heß, S. High Prevalence of Colistin Resistance Genes in German Municipal Wastewater. *Science of The Total Environment* **2019**, *694*, 133454. <https://doi.org/10.1016/j.scitotenv.2019.07.260>.
- (43) Balcazar, J. L. Bacteriophages as Vehicles for Antibiotic Resistance Genes in the Environment. *PLoS Pathogens* **2014**, *10* (7), e1004219. <https://doi.org/10.1371/journal.ppat.1004219>.
- (44) Ross, J.; Topp, E. Abundance of Antibiotic Resistance Genes in Bacteriophage Following Soil Fertilization With Dairy Manure or Municipal Biosolids, and Evidence for Potential Transduction. *Applied and Environmental Microbiology* **2015**. <https://doi.org/10.1128/AEM.02363-15>.
- (45) Dong, P.; Wang, H.; Fang, T.; Wang, Y.; Ye, Q. Assessment of Extracellular Antibiotic Resistance Genes (EARGs) in Typical Environmental Samples and the Transforming Ability of EARG. *Environment International* **2019**, *125*, 90–96. <https://doi.org/10.1016/j.envint.2019.01.050>.
- (46) Wales, A.; Davies, R. Co-Selection of Resistance to Antibiotics, Biocides and Heavy Metals, and Its Relevance to Foodborne Pathogens. *Antibiotics* **2015**, *4* (4), 567–604. <https://doi.org/10.3390/antibiotics4040567>.
- (47) Larsson, D. G. J.; de Pedro, C.; Paxeus, N. Effluent from Drug Manufactures Contains Extremely High Levels of Pharmaceuticals. *Journal of hazardous materials* **2007**, *148* (3), 751–755. <https://doi.org/10.1016/j.jhazmat.2007.07.008>.
- (48) Bengtsson-Palme, J.; Gunnarsson, L.; Larsson, D. G. J. Can Branding and Price of Pharmaceuticals Guide Informed Choices towards Improved Pollution Control during Manufacturing? *Journal of Cleaner Production* **2018**, *171*, 137–146. <https://doi.org/10.1016/j.jclepro.2017.09.247>.
- (49) Giger, W.; Alder, A. C.; Golet, E.; Kohler, H.; McArdell, C.; Molnar, E.; Siegrist, H.; Suter, M. Occurrence and Fate of Antibiotics as Trace Contaminants in Wastewaters, Sewage Sludges, and Surface Waters. *CHIMIA International Journal for Chemistry* **2003**, *57* (9), 485–491.
- (50) Lindberg, R. H.; Björklund, K.; Rendahl, P.; Johansson, M. I.; Tysklind, M.; Andersson, B. A. V. Environmental Risk Assessment of Antibiotics in the Swedish Environment with Emphasis on Sewage Treatment Plants. *Water research* **2007**, *41* (3), 613–619. <https://doi.org/10.1016/j.watres.2006.11.014>.
- (51) Xu, J.; Xu, Y.; Wang, H.; Guo, C.; Qiu, H.; He, Y.; Zhang, Y.; Li, X.; Meng, W. Occurrence of Antibiotics and Antibiotic Resistance Genes in a Sewage Treatment Plant and Its Effluent-Receiving River. *Chemosphere* **2015**, *119*, 1379–1385. <https://doi.org/10.1016/j.chemosphere.2014.02.040>.
- (52) Hausner, M.; Wuertz, S. High Rates of Conjugation in Bacterial Biofilms as Determined by Quantitative in Situ Analysis. *Applied and Environmental Microbiology* **1999**, *65* (8), 3710–3713.
- (53) Wan, Z.; Varshavsky, J.; Teegala, S.; McLawrence, J.; Goddard, N. L. Measuring the Rate of Conjugal Plasmid Transfer in a Bacterial Population Using Quantitative PCR. *Biophys J* **2011**, *101* (1), 237–244. <https://doi.org/10.1016/j.bpj.2011.04.054>.
- (54) Simonsen, L.; Gordon, D. M.; Stewart, F. M.; Levin, B. R. Estimating the Rate of Plasmid Transfer: An End-Point Method. *J Gen Microbiol* **1990**, *136* (11), 2319–2325. <https://doi.org/10.1099/00221287-136-11-2319>.
- (55) Andrup, L.; Andersen, K. A Comparison of the Kinetics of Plasmid Transfer in the Conjugation Systems Encoded by the F Plasmid from Escherichia Coli and Plasmid PCF10 from Enterococcus Faecalis. *Microbiology (Reading)* **1999**, *145* (Pt 8), 2001–2009. <https://doi.org/10.1099/13500872-145-8-2001>.
- (56) Zhong, X.; Driesch, J.; Fox, R.; Top, E. M.; Krone, S. M. On the Meaning and Estimation of Plasmid Transfer Rates for Surface-Associated and Well-Mixed Bacterial Populations. *J Theor Biol* **2012**, *294*, 144–152. <https://doi.org/10.1016/j.jtbi.2011.10.034>.
- (57) Normander, B.; Christensen, B. B.; Molin, S.; Kroer, N. Effect of Bacterial Distribution and Activity on Conjugal Gene Transfer on the Phylloplane of the Bush Bean (*Phaseolus Vulgaris*). *Applied and Environmental Microbiology* **1998**, *64* (5), 1902–1909.
- (58) Suhartono, S.; Savin, M. Conjugal Transmission of Antibiotic-Resistance from Stream Water Escherichia Coli as Related to Number of Sulfamethoxazole but Not Class 1 and 2 Integrase Genes. *Mob Genet Elements* **2016**, *6* (6), e1256851. <https://doi.org/10.1080/2159256X.2016.1256851>.

Additional figures

Fig. S1. Relations between E and other parameters for the main model after 70 years of simulated time.

Fig. S2. Dependency of the different processes on the H parameter for the main model after 70 years of simulated time.

Fig. S3. Dependency of the different processes on the MH parameter for the main model after 70 years of simulated time.

Fig. S4. Dependency of the different processes on the D parameter for the main model after 70 years of simulated time.

Fig. S5. Dependency of the different processes on the MH parameter divided by the H parameter (resulting in an estimated M parameter) for the main model after 70 years of simulated time.

Fig. S6. Valid parameter ranges (A) and process rates (B) for the main model after 70 years of simulated time with S being fixed to 1.

Fig. S7. Differences between how processes depend on E for the main model (A), and the model where S is fixed to 1 (i.e. neutral fitness costs of ARGs, B). In both cases, data is presented at 70 years of simulated time.

Fig. S8. Valid parameter ranges (A) and process rates (B) for the emergence model after 70 years of simulated time.

Table S1. Literature survey: Mobilization

Study	Species	Test system	Events/Day (average)	Remark
Torres et al., 1991 ¹⁹	<i>Enterococcus faecalis</i>	Tn925	1.89E-07	Also includes a HGT step
Manson et al., 2010 ²¹	<i>Enterococcus faecalis</i>	<i>E. faecalis</i> PAI	4.00 x 10 ⁻¹⁰	Also includes a HGT step
Poyart et al., 1995 ²⁰	<i>Enterococcus faecalis</i>	Tn916	1.48 x 10 ⁻⁴	Also includes a HGT step
Poyart et al., 1995 ²⁰	<i>Enterococcus faecalis</i>	Tn916	5.67 x 10 ⁻⁹	Also includes a HGT step
Poyart et al., 1995 ²⁰	<i>Enterococcus faecalis</i>	Tn916	2.40 x 10 ⁻⁹	Also includes a HGT step
Scott et al., 1988 ¹⁸	<i>Bacillus subtilis</i>	Tn916	1.26 x 10 ⁻⁹	Also includes a HGT step
Franke and Clewell, 1981 ¹⁷	<i>Streptococcus faecalis</i>	Tn916	9.95 x 10 ⁻⁹	Also includes a HGT step

Table S2. Literature survey: Conjugation

Study	Species	Test system	Events/Day (average)	Remark
Hausner and Wuertz, 1999 ⁵²	<i>Escherichia coli</i>	pRK415	5.76×10^{-2}	
Wan et al., 2011 ⁵³	<i>Escherichia coli</i>	F-plasmid	2.16×10^2	
Simonsen et al., 1990 ⁵⁴	<i>Escherichia coli</i>	IncFII-plasmid R1	2.40×10^{-10}	
Andrup and Andersen, 1999 ⁵⁵	<i>Escherichia coli</i>	F-plasmid	2.16×10^2	
Andrup and Andersen, 1999 ⁵⁵	<i>Enterococcus sp.</i>	pCF10	4.18×10^2	
Zhong et al., 2012 ⁵⁶	<i>Escherichia coli</i>	pB10	3.12×10^{-2}	
Jutkina et al., 2016 ⁴¹	<i>Escherichia coli</i>	MGEs derived from wastewater	3.84×10^{-5}	Stress: Sub-MIC tetracycline concentration
Normander et al., 1998 ⁵⁷	<i>Pseudomonas putida</i>	TOL-plasmid	4.46×10^{-3}	
Suhartono and Sarvin, 2016 ⁵⁸	<i>Escherichia coli</i>	Plasmids derived from river water	1.60×10^{-3}	Stress: 20 mg/L nalidixic acid and 80 mg/L sulfamethoxazole or 20 mg/L streptomycin and 16 mg/L tetracycline

Table S3. Dependency of the modelled processes on the MH and H parameters

Origin	MH:10 ⁻¹⁵ to 10 ⁻¹³	MH:10 ⁻¹³ to 10 ⁻¹⁰	MH:10 ⁻¹⁰ to 10 ⁻⁷	MH:10 ⁻⁷ to 10 ⁻⁴	MH:10 ⁻⁴ to 10 ⁻²
pathogens in the human microbiome	26.7%	21.7%	0.588%	0.00194%	<0.00001%
non-pathogenic human-associated bacteria	0.00659%	1.56%	36.3%	83.4%	96.0%
chromosomes of environmental bacteria	0.00030%	0.0633%	1.38%	3.65%	2.92%
MGEs in human-associated bacteria	70.1%	73.7%	59.4%	12.4%	1.02%
MGEs in the environment	3.22%	2.98%	2.25%	0.545%	0.0311%
pathogenic bacteria in the environment	<0.00001%	<0.00001%	<0.00001%	<0.00001%	<0.00001%
Origin	H:10 ⁻¹¹ to 10 ⁻⁹	H:10 ⁻⁹ to 10 ⁻⁷	H:10 ⁻⁷ to 10 ⁻⁵	H:10 ⁻⁵ to 10 ⁻³	H:10 ⁻³ to 10 ⁻¹
pathogens in the human microbiome	99.0%	73.3%	3.63%	0.0394%	0.00033%
non-pathogenic human-associated bacteria	0.717%	5.90%	3.07%	0.303%	0.226%
chromosomes of environmental bacteria	0.0326%	0.246%	0.115%	0.0130%	0.00087%
MGEs in human-associated bacteria	0.231%	19.7%	89.8%	95.6%	96.3%
MGEs in the environment	0.0105%	0.822%	3.36%	4.09%	3.72%
pathogenic bacteria in the environment	0.00001%	0.00001%	<0.00001%	<0.00001%	<0.00001%

Bengtsson-Palme et al. (2021)

Modelling antibiotic resistance gene emergence

Additional supplementary material

Code is available from here:

https://microbiology.se/publ/resistance_emergence_model/code.zip

Full results of the pre-existing (main) model:

https://microbiology.se/publ/resistance_emergence_model/preexisting_model.zip

Full results of the emergence model:

https://microbiology.se/publ/resistance_emergence_model/emergence_model.zip