1 Ecology, more than antibiotics consumption, is the major predictor 2 for the global distribution of aminoglycoside-modifying enzymes

3
 4 Léa Pradier¹, Stéphanie Bedhomme¹

5 ¹Centre d'Ecologie Fonctionnelle et Evolutive, CNRS, University of Montpellier, EPHE, IRD,
 6 Montpellier, France

8 Abstract

9
 10 Antibiotics consumption and its abuses have been historically and repeatedly pointed out as the major
 11 driver of antibiotic resistance emergence and propagation. However, several examples show that
 12 resistance may persist despite substantial reductions in antibiotic use, and that other factors are at
 13 stake. Here we study the temporal, spatial, and ecological distribution patterns of aminoglycoside
 14 resistance, by screening more than 160,000 publicly available genomes for 27 clusters of genes
 15 encoding aminoglycoside-modifying enzymes (AMEGs). We find that AMEGs are ubiquitous: about
 16 25% of sequenced bacteria carry AMEGs. These bacteria were sequenced from all the continents and
 17 terrestrial biomes, and belong to a wide number of phyla. By focusing on several European countries
 18 between 1997 and 2018, we show that aminoglycoside consumption has little impact on the
 19 prevalence of AMEG-carrying bacteria, whereas most variation in prevalence is observed among
 20 biomes. We further analyze the resemblance of resistome compositions across biomes: soil, wildlife,
 21 and human samples appear to be central to understand the exchanges of AMEGs between different
 22 ecological contexts. Moreover, the genomic distribution of AMEGs suggests a selection for widening
 23 resistance spectra, mostly driven by mobile genetic elements. Together, these results support the idea
 24 that interventional strategies based on reducing antibiotic use should be complemented by a stronger
 25 control of exchanges, especially between ecosystems.

27 1. Introduction

28 Antibiotic resistance (AR) is a persistent global public health problem that has increased over
 29 the last decades, with resistances spreading faster and faster after antibiotic introduction in clinical use
 30 (Witzany et al., 2020). Though first concerns about infections by antibiotic resistant bacteria (ARBs)
 31 were formulated in the 1940s, the discovery and development of new antibiotics allowed for treatment
 32 substitutions (Podolsky, 2018) during the first decades of antibiotic use. The discovery that AR was
 33 frequently acquired by horizontal gene transfer (HGT, Watanabe, 1963) and the subsequent
 34 emergence of multi-resistant strains led the international health agencies to start raising the issue, at
 35 the end of the 1970s, that antibiotic resistance propagation was threatening to end the antibiotic
 36 golden area and jeopardize the huge progress made in the treatment of bacterial infectious diseases. In
 37 parallel, the discovery and design of new antibiotics had become more and more difficult (Livermore

et al., 2011). This trend persisted despite the bulk of information provided by recent advances in genomics (da Cunha et al., 2019), thus decreasing the hope for potential treatment substitutions. In 2014, the AMR review estimated that 700,000 deaths per year worldwide were attributable to antibiotic resistances, and predicted 10 million deaths per year by 2050 (O'Neill, 2014, but see de Kraker et al., 2016).

Antibiotics consumption and its abuses have been historically and repeatedly pointed out as the major cause of AR genes propagation (O'Neill, 2014; Podolsky, 2018; Schrijver et al., 2018): the frequency of AR increases in bacteria communities under the selective pressure of antibiotics. To fight this threat, most health agencies are thus focusing their policies on sanitation and mostly on a more reasonable use of antibiotics (World Health Organization, 2015). This perception of the factors driving AR spread and its associated policies still prevail nowadays: e.g. a meta-analysis over 243 studies found a positive correlation between antibiotic consumption and presence of AR (Bell et al., 2014). However, even if antibiotic consumption decreases in most developed countries, AR does not always follow the same path: e.g. a drastic reduction in sulfonamide consumption in the United Kingdom during the 1990s was not followed by a decrease in prevalence of sulfonamide-resistant *Escherichia coli*, which actually remained constant (Enne et al., 2004). Antibiotic consumption therefore does not seem to be the only factor maintaining AR in pathogenic bacteria communities.

Several studies have documented specific examples where antibiotic resistances arrived and spread in new geographic locations through human activities such as food trade, tourism travels or immigration (Berndtson, 2020; Nejdat et al., 2021; Van Boeckel et al., 2019). Although the impact of human activities on the circulation and spread of AR genes is now well documented through the accumulation of specific examples, integrative studies identifying large scale trends are lacking and this absence of general view has been pinpointed as a gap in our knowledge of drivers of antimicrobial resistance (Holmes et al., 2016).

Another strongly overlooked factor is ecology, and a growing number of studies has called for a more comprehensive analysis of AR outside of farms and hospitals (see e.g Bengtsson-Palme et al., 2018). Though data in natural ecosystems remain scarce, AR genes (ARGs) have probably always been natural members of the gene pools of environmental microbial communities (D'Costa et al., 2011). Natural ecosystems may contribute to the spread of AR, both as sources and as vectors of propagation (Bengtsson-Palme et al., 2018; Berendonk et al., 2015; Marti et al., 2014). This realization, combined with an increasing access to genomic data, led to bioinformatic studies which goal was to extend our understanding of AR in ecological contexts that are often overlooked: e.g. β -lactam resistance in dairy industry (Pitta et al., 2016), in slaughterhouses (Lavilla Lerma et al., 2014), in wastewater treatment plants (Karkman et al., 2016), or in natural fresh water (Czekalski et al., 2015). However, even if these studies complement our knowledge on the presence, the frequency, the nature, and the circulation of AR in poorly documented environments (see e.g. Cuadrat et al., 2020; Zhang et al., 2020), descriptions of global distribution patterns and analyses of factors underlying them remain scarce.

In this study, we investigate the relative importance of environmental and genomic factors in

shaping the routes of antibiotic spread on a large scale, focusing on aminoglycoside resistance. Aminoglycosides (AGs) are traditional Gram-negative antibiotics that bind to the 30S subunit of prokaryotic ribosomes and thus inhibit protein synthesis (Mingeot-Leclercq et al., 1999). This antibiotic class was first introduced in 1944 with the successful use of streptomycin against tuberculosis. Several years later, other AGs produced by *Streptomyces spp.* were discovered (kanamycin, spectinomycin, tobramycin, neomycin, apramycin) and introduced in clinics. They were followed in the 1970s by a series of new isolates or derivative synthesized compounds (amikacin, netilmicin, isepamicin, dibekacin, arbekacin, see van Hoek et al., 2011). However, the emergence of resistant strains during the following years, combined with the requirement of administration by injection has reduced the use of AGs in therapies (Murray and Murray, 1991). Nowadays, AGs are only used in humans as a second-line or last-resort treatment for Gram-negative bacteria (Garneau-Tsodikova and Labby, 2016) but they remain frequently used in agriculture and veterinary medicine to treat septicemia and infections of the digestive, respiratory, and urinary tracts in a wide variety of domestic animals (European Medicines Agency, 2017).

The three main mechanisms of AG resistance are (Garneau-Tsodikova and Labby, 2016): (i) decrease in drug uptake (through modifications of membrane permeability or of the peri-membrane ion gradient) and/or increase in drug efflux (through efflux pump activation); (ii) drug inactivating enzymes; and (iii) modification of the drug target by point mutations in the genes coding for the proteins of the ribosomal small subunit (Finken et al., 1993; Sander et al., 1995; Toivonen et al., 1999). Among these mechanisms, aminoglycoside-modifying enzymes (AMEs) are a class of inactivating enzymes that catalyze the transfer of chemical groups on specific residues of the AG molecules, leading to a modified drug which poorly binds to its target (Jana and Deb, 2006). AMEs represent the most common mechanism of AG resistance in clinical isolates and are well characterized biochemically (Ramirez and Tolmasky, 2010). The classical nomenclature of AMEs is based on the group they transfer (acetyltransferases, AACs; nucleotidyltransferases, ANTs; and phosphotransferases, APHs), on the residue modified, and on the resistance profile they confer (Ramirez and Tolmasky, 2010). However, AMEs emerged several times during evolution (see e.g. Salipante and Hall, 2003 for AAC(6') enzymes), so biochemical nomenclatures do not reflect the evolutionary history of any class of AMEs. Finally, many AME genes are carried by mobile genetic elements, which give them the potential to be transmitted both vertically and horizontally (Davies, 1983). They represent today a major threat for the treatment of multidrug-resistant bacteria, notably *Mycobacterium tuberculosis* (Labby and Garneau-Tsodikova, 2013).

Through a computational approach, more than 160,000 publicly available genomes were screened to identify the presence of AME-encoding genes (AMEGs) across the phylogeny of Eubacteria. The present study intended (i) to describe the genomic, geographical and ecological distribution of AMEGs; and based on these data, (ii) to quantify the relative contribution of several key factors (geography, ecology, genomic context, human activities) potentially driving the spread of AMEGs.

2. Results

2.1 Aminoglycoside resistance is widespread across geography, ecology, and phylogeny.

160,987 publicly available Eubacteria genomes were screened for the presence of AMEGs. The list of genomes and their metadata are listed in Supplementary information, Tab. S1. Published sequences of AMEs and genes coding for AMEs had been previously grouped in 27 clusters of homologous genes (CHGs), each containing sequences of genes and proteins very likely to share a common ancestor (see Methods).

46,053 AMEGs were detected in 38,523 genomes (i.e. about one quarter of the genomes screened). Their distribution in 27 CHGs is listed in Table 1, and all the gene coordinates are listed in Supplementary information, Tab. S2. Our dataset included 54 phyla across the phylogeny of Eubacteria with 89.6% of genomes belonging to the 3 most represented phyla: Proteobacteria, Firmicutes, and Actinobacteriota. In the same way, though we analyzed genomes from 13,879 species, only 10 species made up 43.6% of the dataset. Therefore, though we found AMEGs in 23 phyla, Proteobacteria, Firmicutes, and Actinobacteria contained 97.2% of detected genes. The frequency of the resistance carrying genomes presented very contrasting patterns across CHGs and phyla (Figure 1). For each CHG, the phylogenetic diversity of the species in which it was detected was evaluated by calculating Faith's distances (Faith, 1992). This reveals that the number of species carrying these genes ranges from 2 to 468 (respectively for CHG 12 with Faith's distance $d_{\text{Faith}}=1.4$ and CHG 8 with $d_{\text{Faith}}=78.4$), but also that for CHGs present in numerous species, these species can belong to a small number of phyla (e.g. CHG 3, $d_{\text{Faith}}=24.5$ for 406 species) or be largely spread across the bacteria phylogenetic tree (e.g. CHG 6.1, $d_{\text{Faith}}=66.7$ for 418 species). Regarding the phyla, some present a high diversity of CHGs (e.g. Proteobacteria, Actinobacteria, Firmicutes I) whereas others only contain a very small number of different CHGs. This diversity of CHG represented across phyla is however largely explained by the differences in the number of genomes sampled for each phylum (Pearson correlation between the number of genomes sampled and the number of CHGs carried for each phylum, $\rho=0.820$, $p=1.7 \cdot 10^{-6}$), which is itself linked to the presence of species of medical interest in some phyla.

The location and biome metadata could be recovered for 45,574 genomes and from these data, it was established that AME genes were present in samples coming from all the continents, with the exception of Antarctica. In most regions, the prevalence of AMEG-carrying bacteria (ACBs) ranged between 20% and 40%. However, over 50% of ACBs in Indonesia, Mexico, and Turkey, while only ca. 10% of ACBs in Japan, Eastern Europe, and Eastern Africa (Figure 2A). A quite high spatial heterogeneity in terms of the proportion of each CHG among the ACBs was also revealed (Figure 2B): CHG 6.1 is over-represented in the Southern hemisphere (Africa, South-East Asia, Oceania, Brazil), CHG27 is over-represented in Canada and Mexico and at very low proportion elsewhere, whereas Western Europe, the United States and Japan have a rather balanced representation of all CHGs. ACBs were identified in samples coming from all the biomes investigated. The vast majority of them come from clinical samples (55.3%), human samples (22.1%), and farm samples (12.3%).

Despite this large bias in ecological distribution, the prevalence of ACBs per biome varies in a relatively narrow range from 9% of bacteria sampled in domestic animals, to nearly 40% of bacteria sampled in humans (Figure 3). The ecological spread of CHGs is strongly correlated to their phylogenetic spread (Pearson correlation between the number of ecosystems and d_{Faith} , $\rho=0.759$, $p=6.8 \cdot 10^{-6}$). Thus, the CHGs with the largest d_{Faith} are ecologically ubiquitous (CHGs 3, 5.1, 6.1, 7, 8, and 27), while 8 other CHGs with limited phylogenetic diversity were found in at most 3 biomes (CHGs 5.2, 6.3, 12, 16, 17.2, 24, 25, and 28). Interestingly, clinical samples were found to carry resistance genes of all CHG except two and these two CHGs were specific to agrosystems only (CHG28) and agrosystems and farms (CHG 6.3). Moreover, there are significant negative correlations between the date of first sampling and the phylogenetic and ecological spreads (Fig. 4A, respectively $\rho=0.712$, $p=4.4 \cdot 10^{-5}$ and $\rho=0.718$, $p=3.6 \cdot 10^{-5}$): all ecologically ubiquitous CHGs were sampled before 1955, while the CHGs that were never sampled before 2000 could be found in at most 4 biomes.

Our dataset includes bacteria sampled between 1885 and 2019. However, the vast majority of them was sampled recently: 96.3% of genomes were sampled after 1990, and 58.9% after 2010. In this dataset, the first occurrence of an ACB dates from 1905 (Fig. 4A), and the prevalence of ACBs remains constant (about 30%) between 1990 and 2019 (Fig. 4B), though the frequencies of individual CHGs varied much more over time in this same time frame (Fig. 4C), with some CHGs having a regularly increasing (e.g. CHG 5.1, 19) or decreasing (e.g. CHG 6.1) frequency over the whole period or showing one-time frequency peak (e.g. CHG11 in 2003 and 2004). The worldwide dynamics shows more a coexistence across time of a diversity of CHG than sequential substitution of one CHG by another. However, though we also observe coexistence across time at continental levels (Supplementary information, Fig. S1), local time trends suggest that some CHGs might be progressively replaced: e.g. the frequency of CHGs 3 and 5.1 has increased a lot to the expense of CHG 6.1 in Europe, Southeastern Asia and Oceania since the 1990s.

2.2 The European distribution of aminoglycoside resistance is driven by ecology and human exchanges

We investigated the potential role of different factors on the distribution of ACBs in Europe: ecology (structuration in biomes), human exchanges (potential ACB importations through immigration and merchandise imports), and aminoglycoside consumption. This analysis was performed on the timeframe 1997-2018 for which antibiotic consumption data by country and by year was available for Europe. We included in the dataset 16 CHGs for which occurrences were detected in at least 30 genomes sampled in Europe during this timeframe. Most CHGs had very distinct distributions over time and space, so they were analyzed separately.

Within Europe, a Matérn spatial autocorrelation structure was kept in all the models: CHGs tend to cluster between close countries and the autocorrelation structure allowed to control for this. A time autoregressive structure was also kept in all the models, which showed positive autocorrelation

for half of CHGs (the others show negative time autocorrelation). Ecology, human exchanges, and antibiotics consumption are kept as explanatory variables for respectively 16 CHGs, 14 CHGs, and 8 CHGs out of 16. The importance of each explanatory variable varies across CHGs, but ecology overall appears to be the most important one: it increased adjusted R^2_{adj} by 15.8% on average (2.1% on average for human exchanges, and 3.0% on average for antibiotics consumption). Interactions between ecology and human exchanges increased R^2_{adj} by 3.4% (for 6 CHGs), and interactions between ecology and antibiotics consumption increased R^2_{adj} by 3.4% (for 3 CHGs) (Figure 5). The effects of all variables and their interactions for individual CHGs are given in Supplementary information, Tab. S3-S18.

The effect of AG consumption is not unidirectional across all the CHGs: when kept in the selected model, its effect is significantly positive for only three CHGs (CHGs 2, 3, and 4), but negative for three others (CHGs 6.1, 17.1, and 19) and nonsignificant for the others. Human exchanges represented in the analysis by our proxy for potential ACBs influx due to trade, has a positive significant effect on the ACBs prevalence for the majority of CHGs (CHGs 4, 6.1, 8, 11, 14, 17.1, 20, and 23), but a negative significant effect for CHGs 3 and 27. Our proxy for potential bacteria influx due to migration has a negative significant effect on the probability to sample ACBs for CHGs 4, 6.1, and 27, and a positive effect for CHG 3. No significant effect of trade and migration were found for other CHGs. Despite the explanatory importance of ecology and the numerous interactions identified, for most CHGs, the probability of sampling ACBs does not significantly differ between most biomes. However, the probability of sampling ACBs in clinical samples is significantly higher for CHGs 1, 3, 5.1, and 19 (respectively significantly lower for CHG 2) than in other biomes. In the same way, CHGs 2, 8, and 27 are significantly less likely to be sampled in farms, while CHGs 3 and 6.1 are significantly more likely to be sampled in humans. CHG 3 is also more likely to be sampled in soil and freshwater, CHG 2 less likely to be sampled in soil, and CHG 6.1 more likely to be sampled in human habitat. We describe the inferred effects of interactions between ecology and antibiotic consumption, and between ecology and human exchanges in Supplementary information, Text S1.

For the 14 CHGs sampled in Europe for which we kept human exchanges as explanatory variables, we computed new models based on the decomposition of these effects: we replaced the two effects beneath human exchanges (i.e. international trade and immigration) in the model by only one effect: the proxy for either bacteria influx due to imports of one of the products considered, or bacteria influx due to immigration. We did not find any variable that was able to explain the distribution of all these CHGs, not even for a majority of them (see Supplementary information, Tab. S19). At most, 4 categories of imports were able to impact the distribution of 5 CHGs with a positive effect (beverages, animal feedstuff, meat, sweet foods), and 9 categories were able to explain the distribution of 4 CHGs with a positive effect (cereals, cork, dairy and eggs, fuels, medical products, human migration, oil seeds, tanning products, miscellaneous edible products). However, the distribution of several CHGs could be explained by one or two categories only: cork for CHG 4; rubber and textiles for CHG 5.1; oil seeds for CHG 6.1; animal feedstuff for CHG 7; human migration and sugar for CHG 8; fuels for

CHG 11; meat and sweet products for CHG 19; and human migration for CHG 27.

We proceeded to the same analyses for worldwide data on the same timeframe, analyzing only the effects of ecology and human exchanges (see Supplementary information, Tab. S20-S37). Antibiotic consumption data could not be accessed at this spatial scale, but the same trend was found for the other factors: ecology is always the most important explanatory factor, and human exchanges are conserved as an explanatory factor for most CHGs (see Supplementary information, Fig. S2).

2.3 The worldwide distribution of aminoglycoside resistance is mainly driven by ecology

Ecology being the most important explanatory factor for the distribution of AMEGs, we investigated the distribution of the different CHGs across ecosystems. Here we approximated an ecosystem as the intersection of a biome and a geographical division (here using the IMAGE24 subdivision from the *rworldmap* library, South, 2011). We described the aminoglycoside resistome composition profile of an ecosystem as the presence or absence of each CHG, in 5-year time frames between 1990 and 2019. For example, in clinical samples of Western Europe, CHG 12 is present over the 2015-2019 period and not before, while CHG 23 is always present except during the 1995-1999 period. We measured differences in resistome composition as Jaccard index, and evaluated their correlation with ecological similarities (being 1 if two ecosystems belong to the same biome, 0 else) and geographical distances between ecosystems. On a global scale, there is no correlation between resistome composition and geography (Mantel test, 9999 permutations, $Z=9.23.10^9$, $p=0.585$) but AMEG compositions depend on ecology (Mantel test, 9999 permutations, $Z=174.9$, $p=1.10^{-4}$). Ecosystems are thus more similar in terms of AMEG content if they are from the same biome than if they are geographically close.

A network approach was adopted to get a deeper understanding of the ecology of the resistome composition. We converted the matrix of resistome composition similarities into a network where each vertex consists of a resistome and edges are weighted by Jaccard indices. A minimum spanning network was built that only retained similarity-based edges that could form maximum similarity paths between resistomes. The resulting network is displayed in Figure 6. Based on the Louvain clustering algorithm, we found three main ecological modules in this network, clustered by their composition in CHGs: (1) clinical, farms, human, and agricultural samples; (2) soil, human habitat, waste, sludge, and freshwater; (3) domestic animals and wild plants and animals. In this network of resistome compositions, wild plants and animals have the highest betweenness centrality (i.e., it is an intermediate in 14 shortest paths in the network), followed by the human habitat (8 shortest paths), soil (7 shortest paths), and human samples (3 shortest paths). All the other biomes have a null betweenness centrality. These four biomes also have the highest closeness centrality in the network. But degree centrality is actually highest for soil (i.e., it shares connections with 7 other biomes), followed by wild plants and animals and human samples (6 connections each), and by agrosystems and farms (5 connections each). Soil, wildlife, and human samples are thus the three biomes for which resistomes are more similar to each other over several time frames.

We then investigated whether resistome resemblance within each of the ecological modules defined above is correlated to phylogenetic similarity in bacteria community composition (defined as the whole set of bacteria sampled within a geographical and ecological unit, and measured with the phylogenetic Sørensen index). After correction for the difference in modules sample sizes, it was established that resistome resemblance is significantly, positively correlated with bacteria community composition similarity for both module 1 (95% CI on Pearson correlation coefficient, $\rho \in [0.075, 0.335]$) and module 2 (95% CI on Pearson correlation coefficient, $\rho \in [0.042, 0.278]$), but no significant correlation was found for module 3 (95% CI on Pearson correlation coefficient, $\rho \in [-0.042, 0.294]$).

2.4 Resistance gene accumulation is under selection for widening resistance spectra

Among 38,523 ACBs, 6246 (~16%) carried more than one AMEG, and the maximum number of AMEG in a genome was eight. We investigated whether and how this accumulation of AMEGs within genomes lead to a broadening of the resistance spectrum. To do so, we used in separate analyses the classifications provided by two published datasets: either one empirically predicting the resistance phenotypes associated to several ARGs (Feldgarden et al., 2019), or another reviewing and compiling all the antibiotics to which most AMEs were reported to confer resistance (Zárate et al., 2018). These two classifications were used because Feldgarden et al. (2019) probably underestimated each resistance spectrum (due to tests on a limited number of antibiotics), while Zárate et al. (2018) probably overestimated them (due to an exhaustive review from heterogeneous sources). The classification by Feldgarden et al (2019) allowed to infer the resistance spectrum of 11,284 genomes from the 12,982 AMEGs they carried, whereas the one by Zarate et al (2018) allowed to infer the resistance spectrum of 26,006 genomes with 31,910 AMEGs. Using the resistance spectra inferred from Feldgarden et al. (2019), two classical functional dissimilarity indices (functional dispersion, FDis, and Rao's quadratic entropy, RaoQE) indicated that the functional diversity increases, i.e. resistance spectrum widens, with the number of resistance genes carried by a genome (FDis, linear regression, $R^2=0.756$, $p<2.2 \cdot 10^{-16}$ and RaoQE, linear regression, $R^2=0.746$, $p<2.2 \cdot 10^{-16}$). More importantly, these increases occur at a higher rate than expected under the hypothesis of a random assortment of AMEGs among genomes, both for FDis (permutation test, 500 perms., $p=1.2 \cdot 10^{-3}$) and for RaoQE (permutation test, 500 perms., $p=6.7 \cdot 10^{-3}$). The same trends were also observed using the resistance spectra inferred from Zárate et al. (2018).

Among the 46,053 AMEGs detected in the initial screen, 15,374 were detected to be associated with MGE conferring intergenomic mobility (plasmids, prophages, or ICEs), and 9640 were detected to be associated with a MGE conferring intragenomic mobility (integrons or transposable elements). 7277 genes were detected to be associated with both at the same time. The proportion of genes associated with MGEs is actually very heterogeneous between CHGs, going from 0% for CHGs 16, 24, and 29 to 94.8%, 95.7% and 100% for CHGs 13, 5.1, and 5.2 (Figure 7). We thus investigated the role of MGEs in the accumulation of AMEGs. FDis and RaoQE increase both

with the number of MGE-associated AMEGs and with the number of AMEGs not associated with MGEs. Yet, the increase rate is significantly higher than expected under the hypothesis of random assortment for genes associated with MGEs conferring intergenomic mobility (permutation test, 500 perms., $p=1.2 \cdot 10^{-3}$ for FDis, $p=1.2 \cdot 10^{-3}$ for RaoQE), whatever the source used to infer resistance spectra. However, for AMEGs associated with MGEs conferring intragenomic mobility or not associated to MGE, the trends differ depending on how the resistance spectra were inferred, and no conclusion can be drawn. The pattern of resistance spectrum widening with the accumulation of AMEGs in a genome is thus stronger and more robust for genes associated with intergenomic mobility MGEs.

Many MGEs are also known to replicate the genes they carry in the same genome. We thus investigated which pairs of AMEGs could be the result of intragenomic replication. The probability for an AMEG to undergo intragenomic replication is at least 10 times higher when associated with MGEs, under the hypothesis of random MGE-gene assortment (permutation test, 500 perms., $p=1.2 \cdot 10^{-3}$). This probability is similar when considering separately MGEs with intergenomic or intragenomic mobility. MGEs therefore promote the presence of several copies of the same resistance gene in a genome and the broadening of the resistance spectrum by the combination of functionally-distant genes.

3. Discussion

3.1 AMEGs show a ubiquitous presence in sequenced bacteria genomes

In this study, we defined and screened 27 clusters of genes encoding resistance to aminoglycosides across more than 160,000 genomes. These genomes were widely spread over the phylogeny of Eubacteria (135 phyla, from Actinobacteriota to Verrucomicrobiota), over ecology (from hospitals to fauna and flora), over geography (from Arctic to Antarctic), and over time (from 1885 to 2019). About one quarter of these bacteria were found to carry a gene known to provide resistance to aminoglycosides. This very high prevalence of aminoglycoside resistance is likely biased by the much higher availability of genomic data for human, clinical, and agriculture isolates. Yet, the lowest prevalence of ARBs we found in a biome was actually 9%, which is consistent with previous studies on genomic (Pal et al., 2015; Zhang et al., 2020) and metagenomic (Zhang et al., 2020) datasets sampled from multiple biomes. A large proportion (ca. 40%) of the ARGs found were potentially mobile, i.e. associated with plasmids, integrons, prophages, ICEs or transposons. A previous study reported <20% of potentially mobile ARGs: 29.4% for ARGs conferring multidrug resistance, 15.5% for β -lactam resistance, and 10.5% for aminoglycoside resistance (Zhang et al., 2020). Yet, it is to note that Zhang et al. (2020) only focused on plasmids and integrons, which may potentially have lead to an underestimation of the frequency of potentially mobile ARGs, e.g. associated with ICEs, phages, or transposable elements.

ACBs are widely spread over space, time, and ecology. We were able to detect AMEGs in all the biomes considered, and all over the world. The prevalence of ACBs ranged from 64% in Turkey

to 0% in Arctic and Antarctic regions. ACBs were found as early as 1905, i.e. nearly 40 years before the first aminoglycoside was isolated and identified as a potential antibacterial therapeutic agent. This is not a surprise as aminoglycosides are naturally produced by bacteria from the *Streptomyces* and *Micromonospora* genera (Durand et al., 2019) and the selection pressure for the evolution of resistance genes existed long before the clinical use of aminoglycosides, as shown for other antibiotic resistance genes (D’Costa et al., 2011).

ACB prevalence strongly increased between the 1940s and the 1980s, but it has plateaued around 30% since the 1990s (Fig. 4B). The phase of increasing prevalence can easily be explained by the discovery and marketing for clinical use of most AGs. Their massive use must have been a strong selective pressure for the emergence and spread of AMEGs. However, AG consumption has stabilized and/or decreased (at least since 1997 in Europe), so the pre-existing selection pressures have remained stable since the 1990s. Besides, most CHGs coexist over long time periods (Fig. 4C), but they are unevenly distributed across space (Fig. 2B). This apparent coexistence might therefore result from the combination of different local dynamics (Supplementary information, Fig. S1). It is to note that two very distant regions could display the same time trends: e.g., the replacement of CHG 6.1 by CHGs 3 and 5.1 in both Europe and Southern Asia and Oceania since the 1990s, or a surge in the frequency of CHG 11 in both Europe and North America during the 2000s. This suggests that local distributions of AMEGs are driven by local parameters, but that local resistomes are also connected at the global scale. Finally, the ecological and phylogenetic distribution of AMEGs varies to a great extent, from ubiquitous CHGs to some being restricted to a few biomes and species. Interestingly, ubiquitous CHGs were first sampled earlier than the others, and the extent of ecological and phylogenetic spread for each CHG is positively correlated to the time since its first sampling occurrence. It is yet impossible to determine the causation order. On the one hand, from two CHGs that emerged at the same date, those with a larger ecological and phylogenetic “niche” may have higher chances to be sampled by chance, so they would have been sampled earlier than ecologically and phylogenetically restricted CHGs. On the other hand, CHGs which emerged earlier (and were thus sampled earlier than the others) may have disposed of more time to spread across several biomes and phyla, whereas recently emerged CHGs would be restricted to the biomes and phyla they have had the time to spread to.

3.2 Antibiotic resistance prevalence is shaped not only by antibiotic use but also by ecology and human exchanges

Reduction of antibiotic consumption and fighting antibiotic pollution have historically been the most common public health recommendation to control or reduce the spread and rise in frequency of antibiotic resistance (see e.g. Mölsted et al., 2008; Sabuncu et al., 2009). The predominance of this recommendation is also driven by the idea, in evolutionary epidemiology, that selection is the major determinant in the emergence, accumulation, and propagation of resistance (Blanquart et al., 2017; Spicknall et al., 2013). This idea is supported by correlative studies linking antibiotic consumption levels and resistance prevalence, mainly in environments where antibiotic consumption is high such

as hospitals or farms (e.g. Goossens et al., 2005).

In this study, one of the goals was to ask whether the impact of antibiotic use on antibiotic resistance genes prevalence could be inferred on a very broad geographical and temporal scale and also across biomes. The local concentrations of antibiotics are likely to strongly vary between biomes, as antibiotics are mainly used in hospitals, outpatients, farms and agrosystems and antibiotics are found as pollutant in other biomes. Data on the amount of antibiotics prescribed in human and veterinary clinics are available but antibiotic pollution data are globally sparse (though local databases have been set up in recent years (see Umweltbundesamt, 2016) and were not available on a sufficiently broad geographical and temporal scale to be integrated in our analysis. However, a large part of the antibiotics consumed are not absorbed by the organism and are excreted, and are thus found in high quantities in urine and feces (Danner et al., 2019; Gothwal and Shashidhar, 2015). Amounts of consumed antibiotics in hospitals and households can thus be found nearly unchanged in effluents (Kümmerer, 2003), and soils could display inhibitory levels of antibiotics due to manuring (Tello et al., 2012). From soil, wastewater treatment plants, and freshwater, indirect contamination can happen towards other biomes (Gothwal and Shashidhar, 2015). Apart from the pollution due to antibiotics manufacturers (Larsson et al., 2018), the presence of man-made antibiotics in the environment is overall highly correlated to antibiotics consumption (see Li et al., 2016 in which the main source of pollution for each antibiotic in a watershed is tracked). Since antibiotic concentrations in different environments are directly correlated with one another, we can assume that the available antibiotic consumption data are acceptable predictors for the antibiotic concentrations in other biomes.

Our analysis shows that between 1997 and 2018 in Europe, antibiotics consumption is not the main factor explaining AMEGs prevalence patterns. For all AMEG CHGs studied, aminoglycoside consumption was only a minor explanatory factor for the distribution of ARBs, with few positive effects and mainly non-directional effects on the probability to sample ARBs. There are even in this dataset examples of low ARB prevalence occurring in countries with high antibiotics consumption: e.g. over the period, the average prevalence of bacteria carrying CHG 6.1 were 1.5% in France and 18.1% in Austria, whereas these two countries consumed respectively 7.6 and 0.3 tons per 1000km² of aminoglycosides. How can we explain the discrepancy between our findings and the previously established correlations between antibiotic consumption and antibiotic resistance prevalence as well as the obvious selective advantage conferred by resistance genes across large range of antibiotic concentrations (Gullberg et al., 2011)? There is a strong contrast between the 1960-1990 period during which most aminoglycosides were discovered and introduced for human consumption and AMEGs frequency increased (Fig. 4B) and the period studied in the present analysis (1997-2018 in Europe) during which no new aminoglycoside has been introduced, all CHGs studied had already emerged since at least 10 years, and aminoglycoside consumption was on average stable in Europe. The period studied is thus likely to be a post-emergence period, in which the relative weight of selection as evolutionary force driving the resistance gene frequency variation is likely to be lower than in the previous period. It is also important to integrate the fact that several mechanisms can allow resistant strains to thrive even under decreasing antibiotic selection pressure. (i) Selective pressures on

AR are multiple and man-made antimicrobial agents are just one of them. Many ARGs evolved as adaptations to antibiotics produced in natural communities, or as exaptations of metabolic enzymes selected for other functions (transferases, efflux pumps, etc., Martinez, 2012). (ii) ARGs can be maintained in populations by hitchhiking on other genes (e.g. resistance to heavy metals, biocides) carried on the same MGE (Pal et al., 2015). (iii) Carrying an ARG is costly only if the gene is expressed, but ARGs can be silenced in nonselective conditions (Wein et al., 2019) and upregulated in selective conditions only (Jiang et al., 2019; Palmer et al., 2018). The results of our integrated study strongly suggest that reducing antibiotics consumption is a necessary but not a sufficient measure to limit the propagation of antibiotic resistances and reduce their prevalence. Complementary measures based on the reduction of other selection pressures (such as heavy metal pollution maintaining ARG frequency should be implemented (Seiler and Berendonk, 2012). Some studies even argue that the spread and fixation of some AR families may no longer depend on the clinical use of antibiotic agents (Baquero et al., 2015).

3.3 Trade and migration matter more

In our dataset, human exchanges explain a significant part of the variation in the prevalence of ACBs for a higher number of CHGs than antibiotics consumption does. Human exchanges were integrated in our models as imports and immigration weighted by the ARGs frequency in the IMAGE24 region of origin (see Fig. 2B). Over the 1997-2018 period, we observed this trend within Europe, and human exchanges also have an important effect worldwide (Supplementary information, Fig S1 and Tab. S20-S37), although at this larger scale it was impossible to compare their effects to the effect of antibiotic consumption. Numerous examples of direct impact of human activities, outside of antibiotic use, on the emergence, retention, and propagation of AR have been documented: ARBs can be carried over continents by plant (Zurfluh et al., 2015) and animal products (Eltaï et al., 2020; Le Hello et al., 2011), exchanged through international trade, as well as immigrants (Nellums et al., 2018) and travelers in general (e.g. Lepelletier et al., 2011). Among these ARBs, several were identified to have caused human infections and epidemics: e.g. the serotype Kentucky ST198 of *Salmonella enterica* (Le Hello et al., 2011); the German Shiga-toxin producing *Escherichia coli* (Buchholz et al., 2011; Rasko et al., 2011). The ARGs carried by imported strains may then spread to local strains and species by horizontal transfer, thus enriching the local resistome or increasing the local frequency of resistant genes. The decomposition of human exchanges in different good categories revealed that no category, either specific (e.g. animal feedstuff imports) or general (e.g. animal products imports), contributes to explain the distribution of all CHGs. This could be due to different CHGs having different geographical origins, to different CHGs being carried by bacteria with different ecological niches, or to contingency in the different transmission pathways. However, imports of meat, beverages, animal feedstuff, and sweet foods explain a significant part of the variation in the distribution of several CHGs. These imports might be responsible for spreading ARGs over large distances. Yet, these factors were identified by correlation only, so whether these imports

are transmission pathways for these CHGs cannot be assessed without direct sampling of such products.

The importance of migration of AR genes, through trade and human travels, as a factor explaining the variation in AR gene frequency, revealed by our analysis suggests that reducing the import of ARBs would be an efficient way of limiting AR gene propagation. Procedures of ARG monitoring in traded goods have been set up but are probably insufficient. For example, as reported for the European Union (Schrijver et al., 2018), the modalities for detection and characterization of ARGs in meat are not harmonized between countries, and only focus on clinically relevant species, thereby ignoring the risk of horizontal transmission to and from other unmonitored strains. However, infections by ARBs are often difficult to connect to consumption of given products, and most recommendations insist on hygiene during food processing and cooking (Capita and Alonso-Calleja, 2013).

3.4 Ecology matters most

Not only did we find a wide diversity of ARGs outside of hospitals and farms, where antibiotics are the most consumed (see Fig. 3), but differences between biomes explained most of the variance when the frequency of ARBs was modeled over time, space and ecology. Across ecosystems (defined as the intersection of a biome and a geographical unit), we found that the resemblance between resistomes depended more on the biomes in which they were sampled, than on their geographical location: geographically distant ecosystems belonging to the same biome have more similar resistomes than geographically close ones belonging to different biomes. This is consistent with previous studies: functional metagenomic selection experiments show drastic differences in the resistance profiles between soil and human gut microbiota isolates (Gibson et al., 2014), and less than 10% of ARGs sampled in wastewater treatment plants can also be found in other environments (Munck et al., 2015). Yet, consistent with Forsberg et al. (2012) who found ARGs shared with 100% identity between soil bacteria and human pathogens, many ARGs were still shared between biomes, including between natural and anthropized biomes.

Additionally, the study of certain biomes seems crucial to understand the spread of AR. Indeed, in the network of resistomes, two of the three most central biomes are “fauna and flora” and “soil”. This could be explained by both antibiotics and ARGs transferred from humans and anthropized habitat to natural ecosystems (e.g. through manuring and fecal pollution, see Karkman et al., 2019; Xie et al., 2018). Natural ecosystems have actually been suggested to constitute reservoirs for ARGs originating from hospitals (see e.g. Baquero et al., 2009; Tripathi and Cytryn, 2017). Also, most natural aminoglycosides are produced by two bacteria genera, *Streptomyces* and *Micromonospora* (Durand et al., 2019), that live mostly in soil and decaying vegetation. It is therefore possible that many of the CHGs first evolved in soil and flora bacteria as defense mechanism against the chemical warfare of these two genera and remained at high frequency because of continuous selection pressure. This could be illustrated by the fact that, in our dataset, even though 19 times more genomes were sampled in hospitals than in soil, only 2.5 times more CHGs had their first documented

occurrence in clinical samples than in soil. If not all CHGs emerged in soil and wildlife, these two biomes appear at least as a hub for the circulation of ARGs between biomes.

As previously suggested, many different biomes may represent sources and/or reservoirs of AR, before it is transmitted to human-associated bacteria (Bengtsson-Palme et al., 2018). It is thus crucial to identify how AR is passed on between biomes. In the biome network, three major modules of resistome were found: (1) ARGs sampled in hospitals, humans, farms, and agrosystems; (2) ARGs sampled in soil, human habitat, waste, and freshwater; and (3) ARGs sampled in domestic and wild animals and plants. Previous metagenomic studies using abundance-based metrics found that resistome composition is correlated to bacteria community composition: either globally (Pehrsson et al., 2016), or in specific biomes: soil (Forsberg et al., 2014), wastewater treatment plants (Ju et al., 2019), human feces (Pehrsson et al., 2016). In agreement with these studies, while using presence-absence-based indices, we found substantial correlation between resistome similarity and phylogenetic beta-diversity in the first (hospitals, humans, farms, agrosystems) and the second modules (soil, human habitat, waste, and freshwater). This correlation indicates that resistome similarity is at least partially due to species composition similarity, and so to the exchange of ARBs between biomes. Indeed, when bacteria disseminate between ecosystems, both community compositions and resistomes tend to converge. Consequently, our results also indicate that the propagation of resistances is not totally explained by gene exchange by horizontal transfer. It could result from multiple factors ranging from dissemination of ARBs: between ecosystems or from common sources, by abiotic dissemination, (Allen et al., 2010), fecal transmission (Karkman et al., 2019), transmission by wastewater (Vaz-Moreira et al., 2014), or transmission through the food chain (Founou et al., 2016). The correlation between resistome and species composition was not found in the second module (domestic animals, flora and fauna), even though we controlled for the lower sampling in this module. This surprising result could actually be due to a higher environmental fragmentation in these biomes, which could result in biased sampling, as documented in metagenomic studies (e.g. Lu et al., 2014). Moreover, previous results showed that higher spatial structures increase the likelihood of HGT events to low-abundance strains (Cairns et al., 2018), that might not be sampled and therefore decrease the likelihood to find a correlation between resistome composition and community composition. The absence of correlation between resistome and species composition leads to formulate the hypothesis that horizontal transfer might play a more important role in ARG propagation in this module than in the other two. This hypothesis could be tested using HGT detection methods such as the one developed in Corel et al. (2018).

3.5 Combination of AMEGs in a genome is shaped by diversifying selection and MGEs movements.

In our dataset, 40% of AMEGs are carried by MGEs. MGE carriage is known to be a strong determinant of the capacity to propagate within and between genome and has been identified as key element for classifying a given antibiotic resistance in the highest risk category (Martínez et al., 2015). The association with MGE frequently takes the form of an embedded structure: for example,

half of transposable elements carrying ARGs were also located on plasmids (Fig. 7). The embedded structures very likely multiply the propagation potential or combine the intra- and inter-genome propagation potential. At a finer scale, the fraction of genes associated to MGE as well as the identity of the associated MGE strongly differ between CHGs, suggesting that these clusters have contrasted propagation probabilities and pathways.

The MGE carriage of resistance genes also means that they can be acquired from different sources and combined within a genome. A second interesting finding is that the combination of AMEGs within genomes is not random: the resistance spectrum widens at a higher rate with the number of AMEGs than expected by chance, which suggests that the combination of AMEGs in a genome is under selection for functional diversification. This broadening of the resistance spectrum is mainly driven by MGEs with inter-genomic mobility (i.e. plasmids, ICEs, and prophages).

Finally, we also found that MGEs increase the likelihood for a genome to carry several copies of the same ARG. It has to be noted here that the copy number accessible from whole genome sequence data is an underestimation of the actual number of copies of the gene for all genomes in which the ARG is carried by a plasmid, because plasmids are usually present in more than one copy. The presence of several copies of a resistance gene, either because of its association with an intra-genomic MGE or its plasmid carriage not only increases the expression level of ARGs (Depardieu et al., 2007; Sandegren and Andersson, 2009), but also allows bacteria to evolve new antibiotic resistance functions on duplicated sequences (San Millan and Maclean, 2017; Sandegren and Andersson, 2009), thus also potentially participating in the functional diversification.

3.6 Limitations

The approach taken here allows to exploit a large amount of publicly available data to gain a broad scale vision of AMEG circulation and propagation. However, it suffers some drawbacks, mainly linked to the fact that it is based on available data and does not result from a dedicated sampling. On the one hand, our screening is likely to underestimate the frequency of resistance carrying genomes because some AMEG families probably remain unknown and AG resistance can be conferred by other resistance mechanisms such as target change, hydrolysis, etc. (Blair et al., 2015). On the other hand, resistance gene carrying genomes do not necessarily produce resistant bacteria. It is indeed nearly impossible to determine with certainty whether each of the genes identified actually confers AG resistance. So far, the genotype-phenotype relationship is poorly understood for ARGs in general (Hughes and Andersson, 2017), and particularly for AMEs: a single amino acid change is likely to change the enzyme's target, and even to remove its resistance function (Feldgarden et al., 2019; Zárate et al., 2018). Besides, some of the genes we identified might actually not be expressed at all, as the presence of cis-regulatory sequences were not screened and their functionality in the genome of residence is difficult to evaluate. However, our screening might still be considered as a decent approximation when working with a high spatio-temporal scale of genomic data.

A second important limitation is that the genomes available are not representative of the species composition of the different biomes and some of the biomes or geographical location are over-

represented. This means that the frequencies of resistant bacteria established in this study are useful for comparisons between biomes, geographical location and time periods but cannot be taken as absolute estimates. Along the same line, the absence of detected resistance in certain places in our data set is likely due to a lack of published genomic data from these places: for example, we did not detect any AMEG in genomes of bacteria from the Antarctic, when metagenomic studies have shown that AG resistance have evolved in polar communities (Perron et al., 2015).

Besides, environmental data were insufficiently standardized. Regarding sampling locations, many of them were unknown in NCBI Biosamples metadata; and others could only be determined to the precision of the country, sometimes by looking at unrelated columns. We thus chose to use the country scale as spatial grain, in order to consider as many of them as possible, but at the cost of geographical precision. Moreover, because sampling can be scarce in many geographical areas, countries outside Europe were grouped in larger entities. This choice of spatial scales, constrained by the metadata available, might have prevented to uncover dynamics that occur at finer spatial scales. In the same way, we had to rely on unstandardized data regarding the categorization of the biomes in which bacteria were sampled. Our categorization of biomes was an attempt to reproduce current categorization of metagenomes, but unlike metagenomic datasets, the ecosystems in which bacteria genomes are sampled are usually poorly described: e.g. the distinction between human samples and clinical samples is often very subtle. Some samples may therefore have been assigned to the wrong biome. And since sampling is greatly biased towards clinical, human, and farms samples, we chose to merge certain classical ecological categories in order to treat the widest diversity of samples. This categorization thus partly differs from other studies.

3.7 Conclusion and perspectives

The present study provides a broad picture of the spatial, temporal and ecological distributions of AMEGs as well as their association with MGEs and reveals contrasted patterns for the different gene families. It additionally establishes that the recent temporal variations of ACBs in Europe are explained first by ecology, second by human exchanges and last by antibiotic consumption. This means that selection by man-made antibiotics is not the only evolutionary force explaining the frequency of ARGs and its variation, such that interventional strategies based on prudent uses of antibiotics for humans, animals, and plants are likely to be a necessary but insufficient way to control and limit the spread of antibiotic resistance. The importance of ecology and human exchanges in shaping the patterns of AMEG prevalence is adding to the growing body of evidences that AR must also be tackled from One Health and Global Health perspectives (Hernando-Amado et al., 2020, 2019). It depends not only on clinical therapeutic guidelines, but also on the high interconnectivity of ecosystems, both locally and globally. Thus, though continuing AR monitoring in clinical and farms samples is crucial, current sampling methods hugely bias genomic datasets and insufficient standardization of data limits their exploitation. Understanding the big picture on AR will require a stronger sampling effort in natural ecosystems and as we have shown that resistomes tend to cluster by ecology rather than by geography, emphasis should be put on monitoring the resistome of

all biomes with equal intensity. It is important to note that metagenomic datasets currently allow neither to precisely assign ARBs to taxonomy, nor to analyze ARG accumulation in genomes, but that their assembly into draft genomes (van der Walt et al., 2017) could help to quickly widen the range of genomic data available from natural ecosystems in the near future.

Finally, this study highlights that AMEGs are frequently associated with MGE but also shows that this level of association strongly varies between gene families. It additionally reveals the role of MGE in the generation of within genome duplications and even more importantly in functional diversification and resistance spectrum broadening. MGEs are known to be vehicles of HGT and are likely to participate in the spread of AMEGs but the strong correlation between resistome composition and species composition established here within biome groups suggests that AMEGs are also spreading by ACBs exchange between biomes. The relative contributions of HGT and ACB migration to AR propagation as well as the factors that shape and orient them should be investigated on datasets as the one generated in this study in order to understand the antibiotic resistance traffic rules and potentially interfere with them to reduce and slow down their movements.

4. Material and methods

4.1 Detection of aminoglycoside modifying enzyme genes

The first aim of this study was to document the geographic, historical and ecological extension of gene families that encode AMEs. This was done by screening for these genes across the clade of Eubacteria. Current methods to detect ARGs in genomic datasets can be, for example, based on HMM profiles drawn from biochemical nomenclatures (e.g. AMRFinder, Feldgarden et al., 2019). However, the biochemical characteristics of AMEs do not reflect their evolutionary history and many profiles may thus gather both homologous and paralogous sequences. Such mixed and evolutionary heterogeneous profiles do not allow to track the spread of a gene family as they can gather sequences with different evolutionary origins despite their similar biochemical functions. New HMM profiles based on clusters of homologous sequences only were thus built. While it may detect a wider diversity of resistance genes, it also increased the need for additional filtering steps.

4.1.1 Definition of resistance profiles

204 sequences of aminoglycoside modifying enzymes were gathered from examples reviewed in Ramirez and Tolmasky (2010) and Garneau-Tsodikova and Labby (2016). From the sequences contained in these two references, two categories have been excluded for further analyses: sequences for which (i) the protein could not be found from the provided identifier solely, and (ii) two enzymatic functions were assigned (e.g. AAC(6')-APH(2'')). All the other sequences (Supplementary information, Tab. S38) were grouped in 29 clusters of homologous genes (CHGs) using SiLiX (v. 1.2.9, Miele et al., 2011) with the following parameter values: at least 70% overlap and 35% identity to accept BLAST hits for building families. For each CHG, sequences were aligned with Clustal Omega (v. 1.2.4, Sievers and Higgins, 2014), and alignments were submitted to HMMER (v. 3.3,

Eddy, 2011) to define profiles. Sequences used to define CHGs, their biochemical classification, and their assignment to CHGs, are listed in Supplementary information, Tab. S38: AACs were split into 14 CHGs, ANT into 6 CHGs, and APH into 9 CHGs, and no CHG gathered proteins with distinct enzymatic activities.

4.1.2 Bacterial genomes

The genomes included in this study are all the Eubacteria genomes available in the NCBI Refseq database on August 18, 2019 and that could be assigned at the species taxonomic level, to the taxonomy *bac120* (release 89, Parks et al., 2018) with the pipeline GTDB-tk (v. 1.0.2, Chaumeil et al., 2020). This pipeline extracts sequences of 120 conserved genes to assign unannotated genomes to the taxonomy, based on the sequence of these 120 genes. 160,987 genomes were used in this study.

4.1.3 Screening resistance profiles into genomes

The genomes were screened for the presence of the 27 CHG profiles defined using the *hmmsearch* command of HMMER (v. 3.3). Around 1,200,000 hits were found in the 160,987 genomes screened. Results were kept with cutoff values of 10^{-3} for e-value and 0.95 for accuracy, and if the length of the predicted proteins was at least 80% of the one of the corresponding profile. When the same protein corresponded to two (or more) overlapping profiles, the result with the lowest e-values was kept. The functions of the proteins screened by HMMER were predicted using InterProScan (v. 5.40, Jones et al., 2014) and proteins with a predicted function incompatible with aminoglycoside modification were filtered out of the dataset based on a keyword list search (see Supplementary information, Text S2 for the list of keywords).

After screening and functional filtering, some CHG profiles did not match any output homologs (i.e. CHGs 9, 10, 21, 22 and 30) and they were thus excluded. The proteins conserved after this functional filtering step were reclustered with SiLiX (with a minimum of 40% identity over 80% overlap) in order to confirm homology. As some sub-clustered CHGs also contained a low number of hits (2 or 3), only clusters that contained at least 10 sequences were kept (<100 sequences excluded). Three CHGs had to be subdivided and the resulting subclusters will be referred to as their initial CHG name, plus another digit (e.g. CHG 6.1). This resulted in 46,053 protein sequences predicted to code for AMEs, belonging to 27 CHGs, and spread across 38,523 genomes.

4.2 The genomic context of aminoglycoside modifying enzyme genes

A search for mobile genetic elements in the neighborhood of predicted AMEGs was performed. The software programs and methods used are detailed thereafter. All programs mentioned below were applied to complete genomes for completely assembled genomes, or to each contig containing AMEGs from partially assembled genomes. The aim was to identify MGEs allowing intra-genomic mobility (transposable elements and integrons) and those allowing inter-genomic mobility (plasmids, prophages, and ICEs). Importantly, about one third of the identified AMEGs were located less than

693 1kb of the extremities of the corresponding contigs, which might have reduced our ability to detect
694 MGEs.

695

696 **4.2.1 Identification of genetic elements allowing inter-genomic mobility**

697 The PlasForest pipeline (v. 1.0.0, Pradier et al., 2021) was used to identify plasmids. Full contigs
698 containing previously identified AMEGs were submitted to this random forest classifier-based
699 software. When a contig was predicted to be of plasmid origin, all the AMEGs it carried were
700 considered as carried by a plasmid. Prophages were then identified with the pipeline PhiSpy (v. 4.0.0,
701 Akhter et al., 2012). AMEGs were considered to be associated with prophages when located within a
702 predicted prophage, or within 1kb upward or downward of it. At last, integrative and conjugative
703 elements (ICEs, sometimes referred to as ‘conjugative transposons’) were identified with the pipeline
704 ICEFinder (Liu et al., 2019). This pipeline can only process annotated genomes, so AMEG-carrying
705 genomes were beforehand annotated with the software Prokka (v. 1.14.5, Seemann, 2014). ARGs
706 were considered to be associated with ICEs when located within a predicted ICE, or within 1kb
707 upward or downward of it.

708

709 **4.2.2 Identification of genetic elements allowing intra-genomic mobility**

710 There are two types of bacterial transposable elements that are likely to carry accessory genes:
711 composite and noncomposite transposons. However, there is currently no standardized method to
712 detect transposable elements in prokaryote genomes (see Goerner-Potvin and Bourque, 2018, with the
713 exception of Red that is not specific to transposable elements, Girgis, 2015). Thus, in order to detect
714 composite transposons, ISfinder database of insertion sequences (Siguier et al., 2006) was
715 downloaded on 2020/05/05 and used as a BLASTn subject database (v. 2.11.0, Camacho et al., 2009).
716 AMEG-carrying contigs were submitted to BLASTn against this database, and only hits larger than
717 minimum IS size (depending on the family, from 600bp for IS200/IS605 to 3000bp for Tn3) were
718 kept. An AMEG was considered as potentially carried by a composite transposon if it was surrounded
719 by at least 2 IS of the same family within 20kb from each other. In order to detect noncomposite
720 transposons, 345,657 transposase protein sequences were extracted from NCBI (list of accession
721 numbers in Supplementary information List. S1) and used as a BLASTp subject database. Contigs
722 carrying AMEGs were translated on 6 frames, then submitted to BLASTp against this database, where
723 only hits larger than 50 amino acids were kept (in order to detect only hits with substantial coverage
724 of transposases). In order to distinguish IS transposases (involved in composite transposon
725 transposition) from other transposases, AMEGs in the neighborhood of IS (previously identified as
726 associated to composite transposons) were excluded from this non-composite transposon screen.
727 Remaining AMEGs were considered as potentially associated with non-composite transposons if they
728 were located within 20kb from a transposase.

729 At last, integrons were identified with the pipeline IntegronFinder (v. 1.5.1, Cury et al., 2016).
730 AMEGs were considered to be associated with integrons when located within a predicted integron, or
731 within 1kb upward or downward of it.

38

4.3 Environmental contexts of ARG-carrying bacteria

4.3.1 Metadata

Through the NCBI BioSamples database, metadata were collected for 90,751 samples (about 55% of the genomes screened for AMEGs).

A classification was attempted for all the samples into the 11 ecological units (biomes) that were defined: clinical environments; freshwater; sludge and waste; wild fauna and flora; domestic animals; farms and farm animals; agrosystems; sea water; human habitat; human samples; soil. To do so, keywords were searched (Supplementary information, Text S3), in a precise order to avoid multiple allocation of the same sample. Some form of sample location was available in the metadata of 75,581 samples: either in the form of geographical coordinates (13,053 samples), location names (exact location, region, or country, 52,898 samples), or institution (university, laboratory, or hospital, 7970 samples). As most locations were not more precise than the country scale, each sampling location was assigned to a country by reverse search in Google Geocode API. Sea water samples could not be assigned to any country, and were thus not included in analysis of spatial variation in AMEGs prevalence. Due to the large number of geographical areas to consider, countries were gathered into larger regions (IMAGE24 classification, R library *rworldmap*, South, 2011). Samples originating from European Union member countries kept their country-level assignment for analyses focusing on Europe only. Sampling dates recovered from the metadata were considered to the time level of year. In total, geographical information, ecological information, and sampling dates could be recovered for 45,574 genomes only.

4.3.2 Antibiotic consumption

Aminoglycoside consumption data in the European Union were collected from the ESAC-NET (for human consumption, European Centre for Disease Prevention and Control, 2021) and the ESVAC (for animal consumption, European Medicines Agency, 2021) databases for the period 1997-2018. Antibiotic consumption data are not freely available for countries outside the European Union. As an aim of this study was to test the impact of antibiotic consumption in general, both sources of antibiotic consumption were summed into a single variable, measured in metric tons of aminoglycosides consumed each year by each country. ESVAC consumption data were measured in metric tons, while ECDC consumption data were reported in Defined Daily Doses (DDDs) for 1000 habitants. ECDC consumption data were thus converted to the same unit using 1 gram as a baseline for aminoglycoside DDD, which is the current WHO standard for amikacin (WHO Collaborating Centre for Drug Statistics Methodology, 2020), and each country's population, recovered from the World Bank (2021a). In order to standardize consumption data between countries of heterogeneous sizes, these measures were divided by each country's surface. Then, missing human and animal consumption data

on the period 1997-2018 were predicted by interpolating the actual data between two consecutive available data points (linear regression: Antibiotic consumption in metric tons per 1000km² ~ Country x Year, R²=0.959 for ESVAC data, R²=0.928 for ECDC data). Finally, these data were summed to a single variable of AG consumption per country and per year.

4.3.3 Human exchanges

The bilateral trade matrix, including imports between 1997 and 2018, was collected from the United Nations Conference on Trade and Development (United Nations Conference for Trade and Development, 2021). This matrix measured imports in 2018 US dollars. From this matrix, imports of energy, information, and services were excluded, to only keep 27 import categories (categories are listed in the rows of Supplementary information, Tab. S19). The bilateral immigration matrix was collected from the World Bank (2021b). Missing immigration data were predicted by interpolating the actual data (linear regression: Number of migrants ~ Country of origin x Country of destination x Year, R²=0.883).

Potential AMEG influx from one country to another was calculated as the sum of imports (respectively immigrants) from the origin region, multiplied by the frequency of AMEGs sampled in the origin region. Total AMEG influx in a country due to trade (respectively immigration) was computed by summing influxes from all possible origins.

4.4 Inference of resistance spectrum

The aim was to assign a resistance spectrum to each of the AMEG we identified. AR phenotypes in particular can be either determined by empirically testing the resistance phenotype associated to a given genotype as has been done in the AMRFinder database (Feldgarden et al., 2019), or inferred from the systematic compilation of all the antibiotics to which a biochemical family is known to resist (Zárate et al., 2018). Here both datasets were used to infer a minimum and maximum resistance spectrum of the AMEGs detected in Eubacteria genomes. An AME biochemical class was attributed to each AMEG sequence identified thanks to the AMRFinder database. A resistance spectrum was then attributed to each AMEG sequence using the established correspondence between the biochemical classes and the AR phenotypes, either empirically predicted (Feldgarden et al., 2019), or reviewed (Zárate et al., 2018). Even if an assignment of each AMEG sequence to an AMRFinder profile could be done, an assignment of each of them to a resistance spectrum was not possible, either because no precise resistance was indicated in Feldgarden et al. (2019), or because the biochemical classes corresponding to AMRFinder profiles were too broad compared to the ones from Zárate et al. (2018) and thus did not allow correct assignment. An AR phenotype could be assigned to 12,982 AMEGs out of 11,284 genomes using empirically tested data, and to 31,910 AMEGs out of 26,006 genomes using reviewed data.

4.5 Statistical analysis

4.5.1 Phylogenetic diversity

To evaluate the phylogenetic range across which each CHG is distributed, two measures summing up the phylogenetic distribution were computed using the function *pd* from the R library *picante*: (i) species richness, i.e. the number of species in which a CHG was found; and (ii) Faith's distance (Faith, 1992), i.e. the sum of the lengths of all the branches of the minimum tree spanning all the bacteria carrying a specific CHG.

4.5.2 Factors influencing the distribution of aminoglycoside resistance

For each CHG, the aim was to explain the local frequencies of AG-resistant bacteria, i.e. the likelihood to sample a genome carrying at least one gene from a given CHG in a given ecosystem at a given date, during the 1997-2018 period. The function *fitme* of R library *spaMM* (Rousset and Ferdy, 2014) was used to compute logistic regressions. The random structure of each null model was selected to minimize conditional AIC, among these hypotheses: no structure; a first-order time-autoregressive process; a Matérn correlation structure based on longitude and latitude of each country's centroid; both a time-autoregressive process and a spatial Matérn correlation structure. Several hypotheses were also considered for the explanatory variables. First, AMEG influx due to trade and AMEG influx due to immigration were highly correlated with each other. In the subsequent statistical analysis, these two variables were thus not considered separately: either both of them were considered to explain the distribution of antibiotic resistance, or none of them were. These two variables were thus considered under the denomination of human exchanges. Then, the following variables were included as fixed effects: ecology (as a categorical variable), AG consumption, human exchanges, and their interactions with ecology (as scaled numerical variables). Including the null model, thirteen different combinations of fixed effects were thus simultaneously considered to explain the distribution of a CHG. For each CHG, the model with the lowest conditional AIC was selected. As antibiotic consumption data were only available for Europe, this model selection was only performed for European data. Due to the temporal variability of sampling in Europe, dates were considered every two years. Another model selection was performed, without considering the antibiotic consumption variable, for worldwide data (with countries being gathered into larger geographical units, see section 4.3.1) with dates being considered every year.

To describe the amount of variance explained by the models and the improvement of the complete models from the null models, adjusted McFadden pseudo- R^2 were computed:

$$R_{Adj}^2 = 1 - \frac{\log(L) - K}{\log(L_0)}$$

where L_0 is the likelihood of the null model, L is the likelihood of the considered model, and K is the number of explanatory variables in the model. To measure the unique contribution of each variable, the difference ΔR_{Adj}^2 was computed, i.e. the difference of adjusted McFadden pseudo- R^2 value measured in the same model without the focal variable.

For 14 CHGs, model selection kept human exchanges as one of the explanatory variables. For these

cases, the effect of the 28 components of the “human exchanges” variable was tested. All these components being highly correlated with each other, “human exchanges” was replaced by only one component at a time. Among the 28 resulting models, only those for which the slope estimates for human exchanges were positive were kept. Finally, models with $AIC < AIC_{min} + 10$ were selected (Burnham and Anderson, 2002).

4.5.3 Distribution of resistomes across geography and ecology

In our dataset, an ecosystem was defined as the intersection of a geographical region and a biome. An aminoglycoside resistome (hereafter simplified as “resistome”) was defined as the set of AMEGs that could be found in an ecosystem. Several CHGs appeared during the time frame considered (1990-2018). Thus, in order for resistomes not to depend on time, CHG presences and absences over several time periods were separately considered (from 1990 to 1994, 1995 to 1999, etc.). For example, CHG 1 presence/absence between 1990 and 1994 and its presence/absence between 1995 and 1999 are considered as different variables for resistome composition comparisons, as if they were two independent CHGs. Similarities between resistomes of different ecosystems were computed using Jaccard index, because it is one of the least sensitive to sampling error (see Schroeder and Jenkins, 2018). This index takes a value of 1 when two ecosystems carry exactly the same CHGs, a value of 0 when two ecosystems carry no common CHG.

To test the dependence of resistomes on geography and ecology, distances were considered between ecosystems (i) either as geographical distances between the centroids of considered regions, or (ii) as binary distances between considered biomes (measured as 1 if two ecosystems belong to the same biome, 0 otherwise). Mantel tests were performed to measure a possible correlation between Jaccard index on one side and the geographical or the biome distance on the other side. Correlation with ecological binary distance was found but not with geographical distance, so ecosystems were then grouped into biomes for further analyses.

To test the association strength between biomes, a network based on the previous matrix of Jaccard indices was created. Each biome was represented as a vertex, and each edge was weighed by the resemblance between two resistomes (measured by Jaccard index). This complete network was pruned into a minimum spanning network (Bandelt et al., 1999), in order to remove the edges that were the least likely to correspond to ARG exchange between two resistomes. Modules were established on this network by using the Louvain method (Blondel et al., 2008). Finally, the relationship between AMEG composition and bacteria community composition was studied at the ecosystem level (each ecosystem being defined as the intersection of a biome and a geographic unit). To do so, the correlation between the resistome Jaccard index and the similarity in bacteria community composition, represented by a phylogenetic Sørensen index (Bryant et al., 2008), was computed.

4.5.4 Functional diversity

Functional dispersion (FDis) et *Rao’s quadratic entropy* (RaoQE) were computed to characterize

functional dissimilarity, in terms of resistance spectrum conferred by different AMEGs contained in the same genome. For this, the R library FD was used (Laliberté et al., 2014). To measure how functional dissimilarity changes with AMEG number per genome, a linear regression of either the square root of FDis or the square root of RaoQ and the logarithm of the number of AMEGs carried by a genome was computed. Permutation tests were performed in order to compare the observed pattern with expectations under the null hypothesis of random assortment of AMEGs among genomes. The gene-genome matrix was permuted 500 times using the function *permatswap* from the R library *vegan*, and FDis and RaoQ were calculated on each genome for each iteration. For each permutation, linear regressions were computed, which yielded a distribution of slopes, simulated under the null hypothesis of random assortment of AMEGs among genomes. Observed slopes were then compared to this distribution, to assess whether AMEG accumulation is compatible with this hypothesis. This analysis was performed for whole genomes, and then distinguishing between AMEGs associated or not associated with MGEs and then further distinguishing between MGEs with intragenomic and intergenomic mobility.

4.5.5 Gene duplications

For each genome that contains at least two AMEGs belonging to the same CHG, the DNA sequences of these genes were recovered, as well as 1kb upstream and downstream, from the NCBI Nucleotide database. Pairwise alignment of these sequences was performed, in order to measure sequence identity between pairs of AMEGs. Because many gene sequences are identical or nearly identical, sequence identity was also computed between pairs of sequences composed of genes and their close genomic contexts. As a conservative threshold, a group of sequences (e.g. a pair, a triplet, or a quadruplet of sequences) was considered to be the result of gene duplication if AMEGs were at least reciprocally 90% identical and their genomic contexts were at least reciprocally 80% identical. Yet, to test the sensitivity of the subsequent analysis to these arbitrary thresholds, other thresholds were used for the identity between AMEGs (80%, 85%, 90% and 95%) and for the identity between genomic contexts (50%, 70%, 80% and 90%). Results did not substantially vary with these thresholds (Supplementary information, Tab. S39-S40). To test if the association of an AMEG sequence with (at least) one MGE makes it more likely to duplicate, the proportions of duplicated groups of AMEGs were computed both among genes associated with MGEs and among genes that were not. Permutations of the AMEGs-MGEs matrix were performed to compute replicates of these proportions under the hypothesis of random physical association between MGEs and AMEGs. The observed odds ratio was then compared to the distribution obtained by permutation.

Acknowledgments

We thank Christelle Leung and Tazzio Tissot for their helpful comments that helped improve the manuscript, and Alexandre Courtiol and Martijn Callens for stimulating scientific and technical discussions. This work was supported by the ERC HGTCODONUSE (ERC-2015-CoG-682819) to

922 SB.

923

924 **Competing interests**

925 The authors declare that they have no competing interests.

926

927 **References**

928

929 Akhter S, Aziz RK, Edwards RA. 2012. PhiSpy: a novel algorithm for finding prophages in bacterial
 930 genomes that combines similarity- and composition-based strategies. *Nucleic Acids Res*
 931 **40**:e126–e126. doi:10.1093/NAR/GKS406

932 Allen HK, Donato J, Wang HH, Cloud-Hansen KA, Davies J, Handelsman J. 2010. Call of the wild:
 933 antibiotic resistance genes in natural environments. *Nat Rev Microbiol* 2010 **8**:251–259.
 934 doi:10.1038/nrmicro2312

935 Bandelt H-J, Forster P, Röhl A. 1999. Median-Joining Networks for Inferring Intraspecific
 936 Phylogenies. *Mol Biol Evol* **16**:37–48.

937 Baquero F, Alvarez-Ortega C, Martinez JL. 2009. Ecology and evolution of antibiotic resistance.
 938 *Environ Microbiol Rep* **1**:469–476. doi:10.1111/J.1758-2229.2009.00053.X

939 Baquero F, Lanza VF, Cantón R, Coque TM. 2015. Public health evolutionary biology of
 940 antimicrobial resistance: priorities for intervention. *Evol Appl* **8**:223–239.
 941 doi:10.1111/EVA.12235

942 Bell BG, Schellevis F, Stobberingh E, Goossens H, Pringle M. 2014. A systematic review and meta-
 943 analysis of the effects of antibiotic consumption on antibiotic resistance. *BMC Infect Dis* **14**:1–
 944 25. doi:10.1186/1471-2334-14-13/TABLES/2

945 Bengtsson-Palme J, Kristiansson E, Larsson DGJ. 2018. Environmental factors influencing the
 946 development and spread of antibiotic resistance. *FEMS Microbiol Rev* **42**:68–80.
 947 doi:10.1093/femsre/fux053

948 Berendonk TU, Manaia CM, Merlin C, Fatta-Kassinos D, Cytryn E, Walsh F, Bürgmann H, Sørum H,
 949 Norström M, Pons MN, Kreuzinger N, Huovinen P, Stefani S, Schwartz T, Kisand V, Baquero
 950 F, Martinez JL. 2015. Tackling antibiotic resistance: the environmental framework. *Nat Rev*
 951 *Microbiol* 2015 **13**:310–317. doi:10.1038/nrmicro3439

952 Berndtson AE. 2020. Increasing Globalization and the Movement of Antimicrobial Resistance
 953 between Countries. *Surg Infect (Larchmt)* **21**:579–585. doi:10.1089/SUR.2020.145

954 Blair JMA, Webber MA, Baylay AJ, Ogbolu DO, Piddock LJV. 2015. Molecular mechanisms of
 955 antibiotic resistance. *Nat Rev Microbiol*. doi:10.1038/nrmicro3380

956 Blanquart F, Lehtinen S, Fraser C. 2017. An evolutionary model to predict the frequency of antibiotic
 957 resistance under seasonal antibiotic use, and an application to *Streptococcus pneumoniae*. *Proc*
 958 *R Soc B Biol Sci* **284**:20170679. doi:10.1098/rspb.2017.0679

959 Blondel VD, Guillaume J-L, Lambiotte R, Lefebvre E. 2008. Fast unfolding of communities in large
 960 networks. *J Stat Mech* **2008**:P10008. doi:10.1088/1742-5468/2008/10/P10008

961 Bryant JA, Lamanna C, Lè Ne Morlon H, Kerkhoff AJ, Enquist BJ, Green JL. 2008. Microbes on
 962 mountainsides: Contrasting elevational patterns of bacterial and plant diversity. *PNAS August*
 963 **12**:11505–11511.

964 Buchholz U, Bernard H, Werber D, Böhmer MM, Remschmidt C, Wilking H, Deleré Y, an der
 965 Heiden M, Adlhoch C, Dreesman J, Ehlers J, Ethelberg S, Faber M, Frank C, Fricke G, Greiner
 966 M, Höhle M, Ivarsson S, Jark U, Kirchner M, Koch J, Krause G, Lubert P, Rosner B, Stark K,

- 967 Kühne M. 2011. German Outbreak of Escherichia coli O104:H4 Associated with Sprouts.
 968 <http://dx.doi.org/10.1056/NEJMoa1106482> **365**:1763–1770. doi:10.1056/NEJMoa1106482
- 969 Cairns J, Ruokolainen L, Hultman J, Tamminen M, Virta M, Hiltunen T. 2018. Ecology determines
 970 how low antibiotic concentration impacts community composition and horizontal transfer of
 971 resistance genes. *Commun Biol* **2018 11** **1**:1–8. doi:10.1038/s42003-018-0041-7
- 972 Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009.
 973 BLAST+: Architecture and applications. *BMC Bioinformatics* **10**:1–9. doi:10.1186/1471-2105-
 974 10-421
- 975 Capita R, Alonso-Calleja C. 2013. Antibiotic-Resistant Bacteria: A Challenge for the Food Industry.
 976 *Crit Rev Food Sci Nutr* **53**:11–48. doi:10.1080/10408398.2010.519837
- 977 Chaumeil PA, Mussig AJ, Hugenholtz P, Parks DH. 2020. GTDB-Tk: a toolkit to classify genomes
 978 with the Genome Taxonomy Database. *Bioinformatics* **36**:1925–1927.
 979 doi:10.1093/BIOINFORMATICS/BTZ848
- 980 Corel E, Méheust R, Watson AK, Mcinerney JO, Lopez P, Baptiste E. 2018. Bipartite Network
 981 Analysis of Gene Sharings in the Microbial World. *Mol Biol Evol* **35**:899–913.
 982 doi:10.1093/MOLBEV/MSY001
- 983 Cuadrat RRC, Sorokina M, Andrade BG, Goris T, Dávila AMR. 2020. Global ocean resistome
 984 revealed: Exploring antibiotic resistance gene abundance and distribution in TARA Oceans
 985 samples. *Gigascience* **9**:1–12. doi:10.1093/GIGASCIENCE/GIAA046
- 986 Cury J, Jové T, Touchon M, Néron B, Rocha EP. 2016. Identification and analysis of integrons and
 987 cassette arrays in bacterial genomes. *Nucleic Acids Res* **44**:4539–4550.
 988 doi:10.1093/NAR/GKW319
- 989 Czekalski N, Sigdel R, Birtel J, Matthews B, Bürgmann H. 2015. Does human activity impact the
 990 natural antibiotic resistance background? Abundance of antibiotic resistance genes in 21 Swiss
 991 lakes. *Environ Int* **81**:45–55. doi:10.1016/J.ENVINT.2015.04.005
- 992 D’Costa VM, King CE, Kalan L, Morar M, Sung WWL, Schwarz C, Froese D, Zazula G, Calmels F,
 993 Debruyne R, Golding GB, Poinar HN, Wright GD. 2011. Antibiotic resistance is ancient.
 994 *Nature*. doi:10.1038/nature10388
- 995 da Cunha BR, Fonseca LP, Calado CRC. 2019. Antibiotic Discovery: Where Have We Come from,
 996 Where Do We Go? *Antibiot 2019, Vol 8, Page 45* **8**:45. doi:10.3390/ANTIBIOTICS8020045
- 997 Danner MC, Robertson A, Behrends V, Reiss J. 2019. Antibiotic pollution in surface fresh waters:
 998 Occurrence and effects. *Sci Total Environ* **664**:793–804.
 999 doi:10.1016/J.SCITOTENV.2019.01.406
- 1000 Davies JE. 1983. Resistance to Aminoglycosides: Mechanisms and Frequency. *Rev Infect Dis*
 1001 **5**:S261–S267. doi:10.1093/CLINIDS/5.SUPPLEMENT_2.S261
- 1002 de Kraker MEA, Stewardson AJ, Harbarth S. 2016. Will 10 Million People Die a Year due to
 1003 Antimicrobial Resistance by 2050? *PLOS Med* **13**:e1002184.
 1004 doi:10.1371/JOURNAL.PMED.1002184
- 1005 Depardieu F, Podglajen I, Leclercq R, Collatz E, Courvalin P. 2007. Modes and Modulations of
 1006 Antibiotic Resistance Gene Expression. *Clin Microbiol Rev* **20**:79. doi:10.1128/CMR.00015-06
- 1007 Durand GA, Raoult D, Dubourg G. 2019. Antibiotic discovery: history, methods and perspectives. *Int*

- 1008 *J Antimicrob Agents* **53**:371–382. doi:10.1016/J.IJANTIMICAG.2018.11.010
- 1009 Eddy SR. 2011. Accelerated Profile HMM Searches. *PLOS Comput Biol* **7**:e1002195.
- 1010 doi:10.1371/JOURNAL.PCBI.1002195
- 1011 Eltai NO, Yassine HM, El-Obeid T, Al-Hadidi SH, Thani AA Al, Alali WQ. 2020. Prevalence of
- 1012 Antibiotic-Resistant Escherichia coli Isolates from Local and Imported Retail Chicken
- 1013 Carcasses. *J Food Prot* **83**:2200–2208. doi:10.4315/JFP-20-113
- 1014 Enne VI, Bennett PM, Livermore DM, Hall LMC. 2004. Enhancement of host fitness by the sul2-
- 1015 coding plasmid p9123 in the absence of selective pressure. *J Antimicrob Chemother* **53**:958–
- 1016 963. doi:10.1093/JAC/DKH217
- 1017 European Centre for Disease Prevention and Control. 2021. Antimicrobial consumption database
- 1018 (ESAC-Net). [https://www.ecdc.europa.eu/en/antimicrobial-consumption/surveillance-and-](https://www.ecdc.europa.eu/en/antimicrobial-consumption/surveillance-and-disease-data/database)
- 1019 [disease-data/database](https://www.ecdc.europa.eu/en/antimicrobial-consumption/surveillance-and-disease-data/database)
- 1020 European Medicines Agency. 2021. European database of sales of veterinary antimicrobial agents.
- 1021 <https://esvacbi.ema.europa.eu/analytics/saw.dll?PortalPages>
- 1022 European Medicines Agency. 2017. Sales of veterinary antimicrobial agents in 31 European countries
- 1023 in 2017.
- 1024 Faith DP. 1992. Conservation evaluation and phylogenetic diversity. *Biol Conserv* **61**:1–10.
- 1025 doi:10.1016/0006-3207(92)91201-3
- 1026 Feldgarden M, Brover V, Haft DH, Prasad AB, Slotta DJ, Tolstoy I, Tyson GH, Zhao S, Hsu C-H,
- 1027 Mcdermott PF, Tadesse DA, Morales C, Simmons M, Tillman G, Wasilenko J, Folster JP,
- 1028 Klimke W. 2019. Validating the AMRFinder Tool and Resistance Gene Database by Using
- 1029 Antimicrobial Resistance Genotype-Phenotype Correlations in a Collection of Isolates.
- 1030 *Antimicrob Agents Chemother* **63**:e00483--19. doi:10.1128/AAC.00483-19
- 1031 Finken M, Kirschner P, Meier A, Wrede A, Bottger' EC. 1993. Molecular basis of streptomycin
- 1032 resistance in Mycobacterium tuberculosis: alterations of the ribosomal protein S12 gene and
- 1033 point mutations within a functional 16S ribosomal RNA pseudoknot. *Mol Microbiol* **9**:1239–
- 1034 1246.
- 1035 Forsberg KJ, Patel S, Gibson MK, Lauber CL, Knight R, Fierer N, Dantas G. 2014. Bacterial
- 1036 phylogeny structures soil resistomes across habitats. *Nat* **509**:612–616.
- 1037 doi:10.1038/nature13377
- 1038 Forsberg KJ, Reyes A, Wang B, Selleck EM, Sommer MOA, Dantas G. 2012. The shared antibiotic
- 1039 resistome of soil bacteria and human pathogens. *Science (80-)* **337**:1107–1111.
- 1040 doi:10.1126/SCIENCE.1220761
- 1041 Founou LL, Founou RC, Essack SY. 2016. Antibiotic resistance in the food chain: A developing
- 1042 country-perspective. *Front Microbiol* **7**:1881. doi:10.3389/FMICB.2016.01881/BIBTEX
- 1043 Garneau-Tsodikova S, Labby KJ. 2016. Mechanisms of Resistance to Aminoglycoside Antibiotics:
- 1044 Overview and Perspectives. *Medchemcomm* **7**:11. doi:10.1039/C5MD00344J
- 1045 Gibson MK, Forsberg KJ, Dantas G. 2014. Improved annotation of antibiotic resistance determinants
- 1046 reveals microbial resistomes cluster by ecology. *ISME J* **9**:207–216.
- 1047 doi:10.1038/ismej.2014.106
- 1048 Girgis HZ. 2015. Red: An intelligent, rapid, accurate tool for detecting repeats de-novo on the

1049 genomic scale. *BMC Bioinformatics* **16**:1–19. doi:10.1186/S12859-015-0654-5/TABLES/4

1050 Goerner-Potvin P, Bourque G. 2018. Computational tools to unmask transposable elements. *Nat Rev*
 1051 *Genet* **2018 1911** **19**:688–704. doi:10.1038/s41576-018-0050-x

1052 Goossens H, Ferech M, Stichele R Vander, Elseviers M. 2005. Outpatient antibiotic use in Europe and
 1053 association with resistance: a cross-national database study. *Lancet* **365**:579–587.
 1054 doi:10.1016/S0140-6736(05)17907-0

1055 Gothwal R, Shashidhar T. 2015. Antibiotic Pollution in the Environment: A Review. *CLEAN – Soil,*
 1056 *Air, Water* **43**:479–489. doi:10.1002/CLEN.201300989

1057 Gullberg E, Cao S, Berg OG, Ilbäck C, Sandegren L, Hughes D, Andersson DI. 2011. Selection of
 1058 Resistant Bacteria at Very Low Antibiotic Concentrations. *PLOS Pathog* **7**:e1002158.
 1059 doi:10.1371/JOURNAL.PPAT.1002158

1060 Hernando-Amado S, Coque TM, Baquero F, Martínez JL. 2020. Antibiotic Resistance: Moving From
 1061 Individual Health Norms to Social Norms in One Health and Global Health. *Front Microbiol*
 1062 **11**:1914. doi:10.3389/FMICB.2020.01914/BIBTEX

1063 Hernando-Amado S, Coque TM, Baquero F, Martínez JL. 2019. Defining and combating antibiotic
 1064 resistance from One Health and Global Health perspectives. *Nat Microbiol* **4**:1432–1442.
 1065 doi:10.1038/s41564-019-0503-9

1066 Holmes AH, Moore LSP, Sundsfjord A, Steinbakk M, Regmi S, Karkey A, Guerin PJ, Piddock LJ V.
 1067 2016. Understanding the mechanisms and drivers of antimicrobial resistance. *Lancet* **387**:176–
 1068 187. doi:10.1016/S0140-6736(15)00473-0

1069 Jana S, Deb JK. 2006. Molecular understanding of aminoglycoside action and resistance. *Appl*
 1070 *Microbiol Biotechnol* **2006 702** **70**:140–150. doi:10.1007/S00253-005-0279-0

1071 Jiang Z, Paudyal N, Xu Y, Deng T, Li F, Pan H, Peng X, He Q, Yue M. 2019. Antibiotic Resistance
 1072 Profiles of Salmonella Recovered From Finishing Pigs and Slaughter Facilities in Henan, China.
 1073 *Front Microbiol* **0**:1513. doi:10.3389/FMICB.2019.01513

1074 Jones P, Binns D, Chang HY, Fraser M, Li W, McAnulla C, McWilliam H, Maslen J, Mitchell A,
 1075 Nuka G, Pesseat S, Quinn AF, Sangrador-Vegas A, Scheremetjew M, Yong SY, Lopez R,
 1076 Hunter S. 2014. InterProScan 5: genome-scale protein function classification. *Bioinformatics*
 1077 **30**:1236–1240. doi:10.1093/BIOINFORMATICS/BTU031

1078 Ju F, Beck K, Yin X, Maccagnan A, McArdell CS, Singer HP, Johnson DR, Zhang T, Bürgmann H.
 1079 2019. Wastewater treatment plant resistomes are shaped by bacterial composition, genetic
 1080 exchange, and upregulated expression in the effluent microbiomes. *ISME J* **13**:346.
 1081 doi:10.1038/S41396-018-0277-8

1082 Karkman A, Johnson TA, Lyra C, Stedtfeld RD, Tamminen M, Tiedje JM, Virta M. 2016. High-
 1083 throughput quantification of antibiotic resistance genes from an urban wastewater treatment
 1084 plant. *FEMS Microbiol Ecol* **92**:14. doi:10.1093/FEMSEC/FIW014

1085 Karkman A, Pärnänen K, Larsson DGJ. 2019. Fecal pollution can explain antibiotic resistance gene
 1086 abundances in anthropogenically impacted environments. *Nat Commun* **2019 101** **10**:1–8.
 1087 doi:10.1038/s41467-018-07992-3

1088 Kümmerer K. 2003. Promoting resistance by the emission of antibiotics from hospitals and
 1089 households into effluent. *Clin Microbiol Infect* **9**:1203–1214. doi:10.1111/J.1469-
 1090 0691.2003.00739.X

- 1091 Labby KJ, Garneau-Tsodikova S. 2013. Strategies to overcome the action of aminoglycoside-
 1092 modifying enzymes for treating resistant bacterial infections. *Future Med Chem* 5:1285–1309.
 1093 doi:10.4155/FMC.13.80
- 1094 Laliberté E, Legendre P, Maintainer BS. 2014. Package “FD” Type Package Title Measuring
 1095 functional diversity (FD) from multiple traits, and other tools for functional ecology.
- 1096 Larsson DGJ, Andremon A, Bengtsson-Palme J, Brandt KK, de Roda Husman AM, Fagerstedt P,
 1097 Fick J, Flach CF, Gaze WH, Kuroda M, Kvint K, Laxminarayan R, Manaia CM, Nielsen KM,
 1098 Plant L, Ploy MC, Segovia C, Simonet P, Smalla K, Snape J, Topp E, van Hengel AJ, Verner-
 1099 Jeffreys DW, Virta MPJ, Wellington EM, Wernersson AS. 2018. Critical knowledge gaps and
 1100 research needs related to the environmental dimensions of antibiotic resistance. *Environ Int*
 1101 **117**:132–138. doi:10.1016/J.ENVINT.2018.04.041
- 1102 Lavilla Lerma L, Benomar N, Sánchez Valenzuela A, Casado Muñoz M del C, Gálvez A, Abriouel H.
 1103 2014. Role of EfrAB efflux pump in biocide tolerance and antibiotic resistance of *Enterococcus*
 1104 *faecalis* and *Enterococcus faecium* isolated from traditional fermented foods and the effect of
 1105 EDTA as EfrAB inhibitor. *Food Microbiol* **44**:249–257. doi:10.1016/J.FM.2014.06.009
- 1106 Le Hello S, Hendriksen RS, Doublet B, Fisher I, Nielsen EM, Whichard JM, Bouchrif B, Fashae K,
 1107 Granier SA, Jourdan-Da Silva N, Cloeckert A, Threlfall EJ, Angulo FJ, Aarestrup FM, Wain J,
 1108 Weill F-X. 2011. International Spread of an Epidemic Population of *Salmonella enterica*
 1109 Serotype Kentucky ST198 Resistant to Ciprofloxacin. *J Infect Dis* **204**:675–684.
 1110 doi:10.1093/INFDIS/JIR409
- 1111 Lepelletier D, Andremon A, Grandbastien B, Group the NW. 2011. Risk of Highly Resistant
 1112 Bacteria Importation from Repatriates and Travelers Hospitalized in Foreign Countries: About
 1113 the French Recommendations to Limit Their Spread. *J Travel Med* **18**:344–351.
 1114 doi:10.1111/J.1708-8305.2011.00547.X
- 1115 Li J, Zhang Haibo, Chen Y, Luo Y, Zhang Hua. 2016. Sources identification of antibiotic pollution
 1116 combining land use information and multivariate statistics. *Environ Monit Assess* 2016 **188**
 1117 **188**:1–12. doi:10.1007/S10661-016-5439-4
- 1118 Liu M, Li X, Xie Y, Bi D, Sun J, Li J, Tai C, Deng Z, Ou HY. 2019. ICEberg 2.0: an updated
 1119 database of bacterial integrative and conjugative elements. *Nucleic Acids Res* **47**:D660–D665.
 1120 doi:10.1093/NAR/GKY1123
- 1121 Livermore DM, Blaser M, Carrs O, Cassell G, Fishman N, Guidos R, Levy S, Powers J, Norrby R,
 1122 Tillotson G, Davies R, Projan S, Dawson M, Monnet D, Keogh-Brown M, Hand K, Garner S,
 1123 Findlay D, Morel C, Wise R, Bax R, Burke F, Chopra I, Czaplewski L, Finch R, Livermore D,
 1124 Piddock LJ V., White T. 2011. Discovery research: the scientific challenge of finding new
 1125 antibiotics. *J Antimicrob Chemother* **66**:1941–1944. doi:10.1093/JAC/DKR262
- 1126 Lu HP, Lai YC, Huang SW, Chen HC, Hsieh CH, Yu HT. 2014. Spatial heterogeneity of gut
 1127 microbiota reveals multiple bacterial communities with distinct characteristics. *Sci Reports* 2014
 1128 **4**:1–8. doi:10.1038/srep06185
- 1129 Marti E, Variatza E, Luis Balcazar J. 2014. The role of aquatic ecosystems as reservoirs of antibiotic
 1130 resistance. *Trends Microbiol* **22**:36–41. doi:10.1016/j.tim.2013.11.001
- 1131 Martinez JL. 2012. Natural Antibiotic Resistance and Contamination by Antibiotic Resistance
 1132 Determinants: The Two Ages in the Evolution of Resistance to Antimicrobials. *Front Microbiol*
 1133 **0**:1. doi:10.3389/FMICB.2012.00001

- 1134 Martínez JL, Coque TM, Baquero F. 2015. Prioritizing risks of antibiotic resistance genes in all
1135 metagenomes. *Nat Rev Microbiol* **13**:396–396. doi:10.1038/nrmicro3399-c2
- 1136 Miele V, Penel S, Duret L. 2011. Ultra-fast sequence clustering from similarity networks with SiLiX.
1137 *BMC Bioinforma* 2011 121 **12**:1–9. doi:10.1186/1471-2105-12-116
- 1138 Mingeot-Leclercq M-P, Glupczynski Y, Tulkens PM. 1999. Aminoglycosides: Activity and
1139 Resistance. *Antimicrob Agents Chemother* **43**:727.
- 1140 Mölsted S, Erntell M, Hanberger H, Melander E, Norman C, Skoog G, Lundborg CS, Söderström A,
1141 Torell E, Cars O. 2008. Sustained reduction of antibiotic use and low bacterial resistance: 10-
1142 year follow-up of the Swedish Strama programme. *Lancet Infect Dis* **8**:125–132.
1143 doi:10.1016/S1473-3099(08)70017-3
- 1144 Munck C, Albertsen M, Telke A, Ellabaan M, Nielsen PH, Sommer MOA. 2015. Limited
1145 dissemination of the wastewater treatment plant core resistome. *Nat Commun* 2015 61 **6**:1–10.
1146 doi:10.1038/ncomms9452
- 1147 Murray BE, Murray B. 1991. New Aspects of Antimicrobial Resistance and the Resulting Therapeutic
1148 Dilemmas. *J Infect Dis* **163**:1185–1194.
- 1149 Nejdat A, Diaz-Reck D, Gelfand I, Zaady E. 2021. Persistence and spread of tetracycline resistance
1150 genes and microbial community variations in the soil of animal corrals in a semi-arid planted
1151 forest. *FEMS Microbiol Ecol* **97**. doi:10.1093/FEMSEC/FIAB106
- 1152 Nellums LB, Thompson H, Holmes A, Castro-Sánchez E, Otter JA, Norredam M, Friedland JS,
1153 Hargreaves S. 2018. Antimicrobial resistance among migrants in Europe: a systematic review
1154 and meta-analysis. *Lancet Infect Dis* **18**:796–811. doi:10.1016/S1473-3099(18)30219-6
- 1155 O'Neill J. 2014. Antimicrobial Resistance: Tackling a crisis for the health and wealth of nations.
1156 *London Rev Antimicrob Resist*.
- 1157 Pal C, Bengtsson-Palme J, Kristiansson E, Larsson DGJ. 2015. Co-occurrence of resistance genes to
1158 antibiotics, biocides and metals reveals novel insights into their co-selection potential. *BMC*
1159 *Genomics* 2015 161 **16**:1–14. doi:10.1186/S12864-015-2153-5
- 1160 Palmer AC, Chait R, Kishony R. 2018. Nonoptimal Gene Expression Creates Latent Potential for
1161 Antibiotic Resistance. *Mol Biol Evol* **35**:2669–2684. doi:10.1093/MOLBEV/MSY163
- 1162 Parks DH, Chuvochina M, Waite DW, Rinke C, Skarshewski A, Chaumeil PA, Hugenholtz P. 2018.
1163 A standardized bacterial taxonomy based on genome phylogeny substantially revises the tree of
1164 life. *Nat Biotechnol* 2018 3610 **36**:996–1004. doi:10.1038/nbt.4229
- 1165 Pehrsson EC, Tsukayama P, Patel S, Mejía-Bautista M, Sosa-Soto G, Navarrete KM, Calderon M,
1166 Cabrera L, Hoyos-Arango W, Bertoli MT, Berg DE, Gilman RH, Dantas G. 2016.
1167 Interconnected microbiomes and resistomes in low-income human habitats. *Nat* 2016 5337602
1168 **533**:212–216. doi:10.1038/nature17672
- 1169 Perron GG, Whyte L, Turnbaugh PJ, Goordial J, Hanage WP, Dantas G, Desai MM. 2015. Functional
1170 Characterization of Bacteria Isolated from Ancient Arctic Soil Exposes Diverse Resistance
1171 Mechanisms to Modern Antibiotics. *PLoS One* **10**:e0069533.
1172 doi:10.1371/JOURNAL.PONE.0069533
- 1173 Pitta DW, Dou Z, Kumar S, Indugu N, Toth JD, Vecchiarelli B, Bhukya B. 2016. Metagenomic
1174 Evidence of the Prevalence and Distribution Patterns of Antimicrobial Resistance Genes in
1175 Dairy Agroecosystems. *Foodborne Pathog Dis* **13**:296–302. doi:10.1089/fpd.2015.2092

- 1176 Podolsky SH. 2018. The evolving response to antibiotic resistance (1945–2018). *Palgrave Commun*
1177 2018 41 4:1–8. doi:10.1057/s41599-018-0181-x
- 1178 Pradier L, Tissot T, Fiston-Lavier A-S, Bedhomme S. 2021. PlasForest: a homology-based random
1179 forest classifier for plasmid detection in genomic datasets. *BMC Bioinforma* 2021 221 22:1–17.
1180 doi:10.1186/S12859-021-04270-W
- 1181 Ramirez MS, Tolmasky ME. 2010. Aminoglycoside modifying enzymes. *Drug Resist Updat* 13:151–
1182 171. doi:10.1016/J.DRUP.2010.08.003
- 1183 Rasko DA, Webster DR, Sahl JW, Bashir A, Boisen N, Scheutz F, Paxinos EE, Sebra R, Chin C-S,
1184 Iliopoulos D, Klammer A, Peluso P, Lee L, Kislyuk AO, Bullard J, Kasarskis A, Wang S, Eid J,
1185 Rank D, Redman JC, Steyert SR, Frimodt-Møller J, Struve C, Petersen AM, Krogfelt KA,
1186 Nataro JP, Schadt EE, Waldor MK. 2011. Origins of the E. coli Strain Causing an Outbreak of
1187 Hemolytic–Uremic Syndrome in Germany. <https://doi.org/10.1056/NEJMoa1106920> 365:709–
1188 717. doi:10.1056/NEJMOA1106920
- 1189 Rousset F, Ferdy J-B. 2014. Testing environmental and genetic effects in the presence of spatial
1190 autocorrelation. Testing environmental and genetic effects in the presence of spatial
1191 autocorrelation. *Ecography (Cop)* 37:781–790. doi:10.1111/ecog.00566i
- 1192 Sabuncu E, David J, Bernède-Bauduin C, Pépin S, Leroy M, Boëlle PY, Watier L, Guillemot D. 2009.
1193 Significant Reduction of Antibiotic Use in the Community after a Nationwide Campaign in
1194 France, 2002–2007. *PLOS Med* 6:e1000084. doi:10.1371/JOURNAL.PMED.1000084
- 1195 Salipante SJ, Hall BG. 2003. Determining the limits of the evolutionary potential of an antibiotic
1196 resistance gene. *Mol Biol Evol* 20:653–659. doi:10.1093/molbev/msg074
- 1197 San Millan A, Maclean RC. 2017. Fitness Costs of Plasmids: a Limit to Plasmid Transmission.
1198 doi:10.1128/microbiolspec.MTBP-0016-2017
- 1199 Sandegren L, Andersson DI. 2009. Bacterial gene amplification: implications for the evolution of
1200 antibiotic resistance. *Nat Rev Microbiol* 2009 78 7:578–588. doi:10.1038/nrmicro2174
- 1201 Sander P, Meier A, Böttger E. 1995. rpsL+: a dominant selectable marker for gene replacement in
1202 mycobacteria. *Mol Microbiol* 16:991–1000. doi:10.1111/J.1365-2958.1995.TB02324.X
- 1203 Schrijver R, Stijntjes M, Rodríguez-Baño J, Tacconelli E, Babu Rajendran N, Voss A. 2018. Review
1204 of antimicrobial resistance surveillance programmes in livestock and meat in EU with focus on
1205 humans. *Clin Microbiol Infect* 24:577–590. doi:10.1016/J.CMI.2017.09.013
- 1206 Schroeder PJ, Jenkins DG. 2018. How robust are popular beta diversity indices to sampling error?
1207 doi:10.1002/ecs2.2100
- 1208 Seemann T. 2014. Prokka: Rapid prokaryotic genome annotation. *Bioinformatics* 30:2068–2069.
1209 doi:10.1093/bioinformatics/btu153
- 1210 Sievers F, Higgins DG. 2014. Clustal Omega. *Curr Protoc Bioinforma* 48:3.13.1-3.13.16.
1211 doi:10.1002/0471250953.BI0313S48
- 1212 Siguier P, Perochon J, Lestrade L, Mahillon J, Chandler M. 2006. ISfinder: the reference centre for
1213 bacterial insertion sequences. *Nucleic Acids Res* 34:D32–D36. doi:10.1093/NAR/GKJ014
- 1214 South A. 2011. rworldmap: A New R package for Mapping Global Data.
- 1215 Spicknall IH, Foxman B, Marrs CF, Eisenberg JNS. 2013. A Modeling Framework for the Evolution
1216 and Spread of Antibiotic Resistance: Literature Review and Model Categorization. *Am J*

- 1217 *Epidemiol* **178**:508–520. doi:10.1093/AJE/KWT017
- 1218 Tello A, Austin B, Telfer TC. 2012. Selective pressure of antibiotic pollution on bacteria of
1219 importance to public health. *Environ Health Perspect* **120**:1100–1106.
1220 doi:10.1289/EHP.1104650
- 1221 Toivonen JM, Boocock MR, Jacobs HT. 1999. Modelling in *Escherichia coli* of mutations in
1222 mitoribosomal protein S12: novel mutant phenotypes of rpsL. *Mol Microbiol* **31**:1735–1746.
1223 doi:10.1046/J.1365-2958.1999.01307.X
- 1224 Tripathi V, Cytryn E. 2017. Impact of anthropogenic activities on the dissemination of antibiotic
1225 resistance across ecological boundaries. *Essays Biochem* **61**:11–21. doi:10.1042/EBC20160054
- 1226 Umweltbundesamt. 2016. Pharmaceuticals in the environment: Global occurrence and potential
1227 cooperative action under the Strategic Approach to International Chemicals Management
1228 (SAICM).
- 1229 United Nations Conference for Trade and Development. 2021. UNCTADstat.
1230 <https://unctadstat.unctad.org/>
- 1231 Van Boeckel TP, Pires J, Silvester R, Zhao C, Song J, Criscuolo NG, Gilbert M, Bonhoeffer S,
1232 Laxminarayan R. 2019. Global trends in antimicrobial resistance in animals in low- And middle-
1233 income countries. *Science (80-)* **365**.
1234 doi:10.1126/SCIENCE.AAW1944/SUPPL_FILE/AAW1944-VAN-BOECKEL-SM.PDF
- 1235 van der Walt AJ, van Goethem MW, Ramond J-B, Makhalanyane TP, Reva O, Cowan DA. 2017.
1236 Assembling metagenomes, one community at a time. *BMC Genomics* 2017 181 **18**:1–13.
1237 doi:10.1186/S12864-017-3918-9
- 1238 van Hoek AHAM, Mevius D, Guerra B, Mullany P, Roberts AP, Aarts HJM. 2011. Acquired
1239 Antibiotic Resistance Genes: An Overview. *Front Microbiol* **0**:203.
1240 doi:10.3389/FMICB.2011.00203
- 1241 Vaz-Moreira I, Nunes OC, Manaia CM. 2014. Bacterial diversity and antibiotic resistance in water
1242 habitats: searching the links with the human microbiome. *FEMS Microbiol Rev* **38**:761–778.
1243 doi:10.1111/1574-6976.12062
- 1244 Watanabe T. 1963. INFECTIVE HEREDITY OF MULTIPLE DRUG RESISTANCE IN
1245 BACTERIA. *Bacteriol Rev* **27**:87–115.
- 1246 Wein T, Hülter NF, Mizrahi I, Dagan T. 2019. Emergence of plasmid stability under non-selective
1247 conditions maintains antibiotic resistance. *Nat Commun* 2019 101 **10**:1–13. doi:10.1038/s41467-
1248 019-10600-7
- 1249 WHO Collaborating Centre for Drug Statistics Methodology. 2020. WHOC - ATC/DDD Index.
1250 https://www.whocc.no/atc_ddd_index/?code=J01GB06
- 1251 Witzany C, Bonhoeffer S, Rolff J. 2020. Is antimicrobial resistance evolution accelerating? *PLoS*
1252 *Pathog* **16**. doi:10.1371/JOURNAL.PPAT.1008905
- 1253 World Bank. 2021a. World Bank Open Data.
- 1254 World Bank. 2021b. Global Bilateral Migration Database.
- 1255 World Health Organization. 2015. Global action plan on antimicrobial resistance.
- 1256 Xie WY, Shen Q, Zhao FJ. 2018. Antibiotics and antibiotic resistance from animal manures to soil: a

1257 review. *Eur J Soil Sci* **69**:181–195. doi:10.1111/EJSS.12494

1258 Zárte SG, Claire MLD la C, Benito-Arenas R, Revuelta J, Santana AG, Bastida A. 2018.

1259 Overcoming Aminoglycoside Enzymatic Resistance: Design of Novel Antibiotics and Inhibitors.

1260 *Mol* 2018, Vol 23, Page 284 **23**:284. doi:10.3390/MOLECULES23020284

1261 Zhang AN, Hou C-J, Negi M, Li L-G, Zhang T. 2020. Online searching platform for the antibiotic

1262 resistome in bacterial tree of life and global habitats. *FEMS Microbiol Ecol* **96**:107.

1263 doi:10.1093/FEMSEC/FIAA107

1264 Zurfluh K, Nüesch-Inderbinnen M, Morach M, Berner AZ, Hächler H, Stephan R. 2015. Extended-

1265 Spectrum-Lactamase-Producing Enterobacteriaceae Isolated from Vegetables Imported from the

1266 Dominican Republic, India, Thailand, and Vietnam. doi:10.1128/AEM.00258-15

1267

Figure 1: Prevalence of AMEG carrying bacteria across the phylogeny of Eubacteria. The phylogenetic tree corresponds to an aggregation of *bac120* phylogeny (Parks et al. 2018) to the phylum level. In the heatmap, blank boxes correspond to the observed absence of a CHG in a phylum. For the other colors, blue to red boxes stand for CHG frequencies from near-zero to one. Gray bars in the top part correspond to the Faith's distance, i.e. the sum of the lengths of all the branches leading to species containing genomes carrying each CHG. Gray bars on the right correspond to the prevalence of ACBs for each phylum, i.e. the proportion of genomes in which at least one AMEG was found.

Figure 2: Distribution of aminoglycoside-resistant bacteria and AME-encoding genes over the world. (A) Distribution of sampled ACBs in the world. The frequencies of resistant and sensitive bacteria are displayed in red and blue, and the size of pies represents the number of genomes sampled in a given region. (B) Distribution of sampled AMEGs in the world. Pie size is irrelevant to the number of samples.

Figure 3: Prevalence of AMEG carrying bacteria across land biomes. In the heatmap, blank boxes correspond to the observed absence of a CHG. For the other colors, blue to red boxes stand for CHG frequencies from near-zero to 0.3. Gray bars on the right correspond to the prevalence of ACBs for each biome, i.e. the prevalence of genomes in which at least one AMEG was found.

Figure 4: Global time trends for aminoglycoside resistance. (A) First sampling occurrence for each CHG in the analyzed dataset. Color points represent the biome in which each CHG was first sampled. CHGs with several color points were sampled in several biomes the first year. (B) Evolution of the worldwide and European prevalences of ACBs between 1943 and 2019. The dots and error bars represent the measured prevalence ($\pm$ standard error) of ACBs each year. The dotted line represents a binomial regression of the prevalence of ACBs over time, fitted separately before 1995 and after 1995. The blue curve represents the number of bacteria genomes sampled each year (in logarithmic scale). The red rectangle displays the period of time analyzed in panel: (C) Evolution of worldwide CHG frequencies among sampled AMEGs between 1990 and 2019. 17 CHGs for which less than 400 sequences were sampled are grouped and displayed in gray.

Figure 5: Relative importance of several factors to explain the distribution of aminoglycoside-resistant bacteria in Europe between 1997 and 2018. Logistic regressions with a spatial Matérn correlation structure were computed to explain the frequency of 16 CHGs. This figure represents the contribution of each variable in each selected model, as the fraction of adjusted McFadden's pseudo- R^2 explained by adding this variable.

1309

1310 **Figure 6: Minimum spanning network based on shared antibiotic resistance genes between**
 1311 **biomes.**

1312

1313

Figure 7: Distribution of the frequency of genes carried by MGEs per CHG. CHGs are organized by descending total frequencies of genes associated with MGEs. The number of genes identified for each cluster is indicated above each bar.

1319 **Table 1: List of CHGs by biochemical function**

Cluster	Biochemical function	Number of genes identified
CHG 1	N-acetyltransferases	619
CHG 2	N-acetyltransferases	711
CHG 3	N-acetyltransferases	4636
CHG 4	N-acetyltransferases	1044
CHG 5.1	N-acetyltransferases	5230
CHG 5.2	N-acetyltransferases	13
CHG 6.1	N-acetyltransferases	11,227
CHG 6.3	N-acetyltransferases	10
CHG 7	N-acetyltransferases	4016
CHG 8	nucleotidyltransferases	4508
CHG 11	nucleotidyltransferases	2990
CHG 12	nucleotidyltransferases	21
CHG 13	nucleotidyltransferases	982
CHG 14	phosphotransferases	675
CHG 15	phosphotransferases	351
CHG 16	phosphotransferases	49
CHG 17.1	phosphotransferases	518
CHG 17.2	phosphotransferases	67
CHG 19	N-acetyltransferases	3124
CHG 20	N-acetyltransferases	871
CHG 23	N-acetyltransferases	2246
CHG 24	N-acetyltransferases	50
CHG 25	nucleotidyltransferases	59
CHG 26	phosphotransferases	20
CHG 27	phosphotransferases	1970
CHG 28	phosphotransferases	23
CHG 29	phosphotransferases	23

1320

1321