## RESEARCH



# Global analysis of the metaplasmidome: ecological drivers and spread of antibiotic resistance genes across ecosystems



## Didier Debroas<sup>1\*</sup>

### Abstract

**Background** Plasmids act as vehicles for the rapid spread of antibiotic resistance genes (ARGs). However, few studies of the resistome at the community level distinguish between ARGs carried by mobile genetic elements and those carried by chromosomes, and these studies have been limited to a few ecosystems. This is the first study to focus on ARGs carried by the metaplasmidome on a global scale.

**Results** This study shows that only a small fraction of the plasmids reconstructed from 27 ecosystems representing 9 biomes are catalogued in public databases. The abundance of ARGs harboured by the metaplasmidome was significantly explained by bacterial richness. Few plasmids with or without ARGs were shared between ecosystems or biomes, suggesting that plasmid distribution on a global scale is mainly driven by ecology rather than geography. The network linking plasmids to their hosts shows that these mobile elements have thus been shared between bacteria across geographically distant environmental niches. However, certain plasmids carrying ARGs involved in human health were identified as being shared between multiple ecosystems and hosted by a wide variety of hosts. Some of these mobile elements, identified as keystone plasmids, were characterised by an enrichment in antibiotic resistance genes (ARGs) and CAS-CRISPR components which may explain their ecological success. The ARGs accounted for 9.2% of the recent horizontal transfers between bacteria and plasmids.

**Conclusions** By comprehensively analysing the plasmidome content of ecosystems, some key habitats have emerged as particularly important for monitoring the spread of ARGs in relation to human health. Of particular note is the potential for air to act as a vector for long-distance transport of ARGs and accessory genes across ecosystems and continents.

Keywords Resistome, Metaplasmidome, One Health

### Background

Microorganisms, the invisible architects of Earth's ecosystems, use an astonishing array of genetic tools to navigate the ever-changing ecosystems they inhabit. Central to their genetic repertoire are plasmids, circular or linear

\*Correspondence:

Didier Debroas

didier.debroas@uca.fr

fragments of DNA that exist alongside chromosomal DNA, carry genes that confer diverse benefits to their host microorganisms [1].

Most plasmids encode accessory genes that can expand the ecological niche of their hosts by, for example, breaking down toxic compounds or providing new metabolic capabilities [2]. These advantages range from metabolic pathways that enable the use of specific resources to virulence factors that aid in pathogenicity [3]. Plasmids are also critical players in the spread of resistance to antibiotics such as penicillins,



© The Author(s) 2025. **Open Access** This article is licensed under a Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License, which permits any non-commercial use, sharing, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if you modified the licensed material. You do not have permission under this licence to share adapted material derived from this article or parts of it. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit http://creativecommons.org/licenses/by-nc-nd/4.0/.

<sup>&</sup>lt;sup>1</sup> Université Clermont Auvergne, CNRS, LMGE, F-63000 Clermont–Ferrand, France

aminoglycosides, sulphonamides and last-resort antibiotics such as colistin or carbapenems [4]. In particular, the spread of antibiotic resistance genes via plasmids is a pressing concern, as most antibiotic resistance genes of public health importance are located in mobile genetic elements (MGEs) [5] including viruses [6]. Unfortunately, most studies of the resistome do not distinguish between antibiotic resistance genes (ARGs) harboured by extrachomosomal elements and chromosomes, whereas these genes may perform functions unrelated to antibiotic resistance [7]. Indeed, many ARGs have been shown to confer resistance only when inserted into MGEs [8]. The associations between ARGs and MGEs are therefore significant and have profound effects on phenotypic resistance [7, 9, 10]. The genetic context, especially as harboured by a plasmid, is thus essential for tackling antibiotic [11]. In addition, the plasmids can include other mobile elements such as insertion sequences (IS), composite transposons, integrons (attC sites and intI genes), integrative and conjugative elements (ICE) [12].

Plasmids therefore act as vehicles for the rapid dissemination of ARGs, contributing to the global challenge of antibiotic resistance whose the success is based on three aspects: the communication space, the vehicle for communication and the interpretation of the message by the recipient [13]. Understanding how plasmidome gene contents interact within environments is pivotal for deciphering the mechanisms underlying the spread of ARGs and devising strategies to combat antibiotic resistance. Intermicrobiome communication can be facilitated by shuttle bacterial species belonging to generalist taxa, able to multiply in the microbiomes of various hosts. Plasmids that can replicate in a broad range of hosts are of concern as they may drive gene exchange over large phylogenetic distances [14]. These broad host-range (BHR) plasmids associated with ARGs have often been isolated from habitats such as produce, soils, manure, wastewater or rivers (cited in Castañeda-Barba et al. [4]). More generally, the gene exchanges could be shaped principally by ecology rather than geography or phylogeny [15]. Nonetheless, it is important to note that these conclusions are primarily drawn from a limited set of ecosystems, with a predominant focus on the human microbiome. To effectively combat the growing threat of antibiotic resistance, it is crucial that current antibiotic resistance prediction procedures are significantly enhanced by considering horizontal gene transfer. Such considerations should be incorporated early in the preclinical analysis of antibiotics. Moreover, when assessing the risk associated with a new antibiotic, it would also be prudent to base this assessment on a comprehensive understanding of the resistome within specific environments where human pathogens could interact with potential bacterial donors [16].

A holistic approach to ARG development and risk assessment has the potential to strengthen our efforts in addressing the challenge of antibiotic resistance. As our understanding of the dynamic relationship between microorganisms and their environments deepens, the study of plasmidome content within structured habitats emerges not only as a scientific endeavour but also as a crucial pursuit within the One Health framework. One Health approach to antibiotic resistance depends on the progress of understanding multi-hierarchical systems, encompassing communications among environments. Although laboratory studies with pure cultures provide invaluable knowledge, it is now important to also incorporate experimental set-ups that aim to reflect the diversity of plasmids. The advent of metagenomic techniques has revolutionised the study of plasmidomes [17]. This approach allows researchers to uncover unexplored genetic reservoirs harboured by plasmids in complex communities (i.e. metaplasmidome). As a result, specific experiments have explored the metaplasmidome in relatively restricted environments using specialised experimental protocols [18]. However, the advent of new procedures that combine high-throughput sequencing and advanced bioinformatics tools now allows the comprehensive decoding of plasmid content in a wide array of metagenomic experiments providing a holistic view of the gene exchanges [19]. This approach not only offers the potential to study the plasmidome on a global scale but also offers insights into the complex interactions between humans and different environmental contexts (wastewater, natural ecosystems...) within the One Health paradigm.

This study, by reconstructing plasmids from various public databases covering different ecosystems (human, insect, aquatic, soil, etc.), represents the first step towards deciphering their genetic content and their involvement in gene flow on a global scale.

#### Results

#### Main characteristics of the metaplasmidome

A total of 16,836,376 plasmid-like sequences (PLSs) were identified from a dataset of 15,023 metagenomes (totalling 985.1 Gb of assembled data). On average, these contigs represented 10.7% of the assembled data in terms of base pairs (bp). After read recruitment, an average plasmid content of 11.3% was observed. However, significant variation was observed, with the proportion of this MGE being as low as 1% in the marine metagenome and rising to 25.1% in the human gut (Supplementary Fig. 1). The human oral and gut, and wastewater treatment plant (WWTP) ecosystems were characterised by higher proportions of plasmid content. The detection of the *att*C cassette recombination sites (expressed as sites per megabase) in the various ecosystems highlights that overall the freshwater ecosystems harboured the highest ratio (>0.44 sites per megabase) (Supplementary Fig. 2). The air metagenomes were also characterised by high ratio of 0.42 *attC* sites per megabase.

After clustering, 6,244,208 non-redundant plasmid-like clusters (PLCs) were found in metagenomes with a mean length of 3.9 kb (max: 868.3 kb). These PLCs constitute the metaplasmidome.

Only a small fraction of PLCs was catalogued in the plasmid RefSeq database (0.1%) and the new IMG/PR database (3.1%), which collects plasmids in microorganisms and metagenomes from JGI. Kmer-based analysis revealed that plasmidomes from various environments and reference databases (such as RefSeq and PLSDB) are clustered together (Supplementary Fig. 3). In particular, metaplasmidomes from the human gut exhibit the closest association with the reference database mMGE, which focuses primarily on human mobile genetic elements and the plasmids [20]. Some PLCs found in ice, human vagina and invertebrate guts are highly divergent from each other. Among the 159,635 PLCs found within the metagenome-assembled genomes (MAGs) of the Earth microbiome catalogue [21], 55,536 were unique.

Out of the PLCs, a total of 25,124,994 proteins were predicted. Among these, 10,879,720 were successfully annotated using the PFAM database, and 2,138,773 matched against the KEGG database. The annotation constructed using the KEGG database was very similar to that obtained using the plasmid reference database (i.e. RefSeq), with a predominance of transposases (Supplementary Fig. 4 and Supplementary Table 1). Overall, compared to chromosomes, the metaplasmidome was significantly enriched in transposases, conjugal transfer pilus and type IV system (*P*-adj < 0.05—DESeq2, Supplementary Fig. 4).

The functional annotation of PLCs with KEGG database enables the visualisation of three main clusters (Supplementary Fig. 5) which share similar types of hallmark and accessory genes. The first cluster encompassed air, human skin, urban and sheep ecosystems. The second cluster predominantly included terrestrial and freshwater environments. The last cluster comprised all animal and human ecosystems, with a few exceptions, such as the sheep ecosystem mentioned earlier, and freshwater sediment and ice ecosystems, which grouped with this third cluster.

#### Resistance genes encoded by PLCs

The ARGs represented 2.44% of the annotated genes from metaplasmidomes. The main types of resistance

identified in the predictions were ABC transporters (33.7%) and glycopeptide resistances (32.6%) (Supplementary Table 2). Mapping metagenomic reads against these antibiotic resistance genes (ARGs) allowed a quantitative exploration of the underlying resistance mechanisms by ecosystem (Fig. 1A) and their distribution worldwide. In general, ARGs were most abundant in human gut and wastewater ecosystems, and notably less frequent in marine environments, regardless of location (Supplementary Fig. 6). This approach allows to group closely related environments based on their predominant resistance type. Interestingly, human and animal guts showed clustering tendencies with the wastewater environment (Fig. 1B). The air, sediment and river metagenomes did not cluster with any of the other environments. The air environment was distinguished by the dominance of the MFS transporter and the river environment by the enrichment of chloramphenicol resistances. There was not relationship between ARGs and metal resistance genes (MRGs) expressed in RPKM (r = 0.082). The abundance of MRGs was highest in the riverine metagenome, followed by wastewater (Supplementary Fig. 7). Among the human biomes, the skin and oral ecosystems showed the highest abundances.

## Mobile genetic elements and CRISPR systems detected in PLCs

A total of 32,895 PLCs carrying integrons (0.5% of PLCs) were detected when searching for attC sites and intI genes. Furthermore, it was observed that PLCs containing antibiotic resistance genes (ARGs) were significantly enriched for integron components (chi-square test, P value < 0.001). Other MGEs capable of integrating into the host DNA were mainly insertion sequences (IS) with terminal inverted repeats (left and right) with 17,982 hits distributed among 16,992 PLCs (0.27% of PLCs). The top 5 of IS identified were ISKpn14, IS401, ISSsp2, IS26 and ISPme1. Among the top 10 of the CDS flanking these MGEs, the majority were hypothetical proteins (94.26%), with few identified as ARGs (1.46%). Other features detected in the PLCs allow the study of interplasmid competition. Specifically, 0.51% of the PLCs encoded CAS proteins, and class 1 systems were found to dominate the metaplasmidome landscape, accounting for 93.1% of the systems. In addition, in the metaplasmidome, 287,730 spacers were detected on 13,937 PLCs targeting 503,786 other PLCs.

#### Main factors associated to metaplasmidome distribution

Overall, there are only a few PLCs that are shared between the different biomes investigated (Supplementary Fig. 8A). The NMDS revealed minimal overlap between the environments (Fig. 2A); the dissimilarity



Fig. 1 A ARG abundances harboured by metaplasmidome (PLCs) in the different ecosystems studied and **B** heat map of the resistance categories (colour is graded from yellow to red to reflect increasing abundance). The quantification was assessed after mapping reads against genes predicted on PLCs (Supplementary Fig. 15D)



**Fig. 2 A** NMDS computed from the distances between all PLCs highlighting a clustering by main biomes. At the level of the ecosystem, the ANOSIM test gives R = 0.92 (P < 0.001). **B** Intersection between the different biomes restricted to PLCs hosting ARGs

between samples was significantly explained by the ecosystem to which they belonged to (ANOSIM statistic r = 0.92, P < 0.01). The main factor driving this clustering pattern based on ecosystem and biome is the microbial community (supplementary material Fig. 8A), dissimilarity and taxonomic diversity, while geographic distance is a weaker predictor (r=0.25, P<0.01 Mantel test). A similar pattern was observed when examining the distances between samples by focusing only on the PLCs with ARGs. These distances were highly correlated with the ecosystem (r = 0.93, P < 0.01, ANOSIM test) and taxonomic composition (r = 0.82, P < 0.01), while showing little dependence on geographic location (r = 0.27, P < 0.01). The curated human gut microbiome database [22] enables to focus on the human microbiome and specifically examine the distribution of PLCs, with particular emphasis on disease markers. Among the potential explanatory variables (countries, sex, disease, body mass index or BMI, and age), BMI was the main factor accounting for the differences between PLCs (P<0.01, PERMANOVA test), although it explained only a small proportion of the total variation in the data (15%). The other variables accounted for only 12.3% of this variation. Ultimately, the distribution of PLCs in the human gut was not effectively explained by these variables alone.

Thus, ecosystems and their associated microbial community composition emerged as the dominant factor influencing plasmid diversity, with some PLCs being shared across multiple habitats. The upset diagram clearly shows that the most common intersections were found between only two biomes: human and wastewater, freshwater and terrestrial or wastewater (Supplementary Fig. 9A). Notably, a limited number of PLCs were found to be shared among by at least three biomes. The most significant cases included human biome in combination with animals, urban and wastewater. In this study, the PLCs affiliated to pBi143, identified as a marker of human faecal contamination [23], were detected not only in the human gut but also in wastewater, air, and, surprisingly, predominantly in riverine ecosystems. These specific PLCs were not detected in the other ecosystems studied. The pattern is different when only the few PLCs referenced in public databases are considered (Supplementary Fig. 9B). In this scenario, many PLCs were shared between different biomes. Focusing solely on PLCs containing ARGs, the most frequent associations were observed between two biomes, specifically freshwater and terrestrial or wastewater (Fig. 2B). The human biome was associated to wastewater and a noteworthy association link this biome to freshwater, wastewater, urban, air and animals. Among animals, pig, chicken and bovine shared PLCs with human gut. Only 308 and 52 PLCs were common to the 8 and 9 biomes, respectively.

#### **Bacteria-PLCs connections**

The presence of protospacers in PLCs or their detection in MAGs allowed the construction of a bipartite network, linking the MAGs constructed by Navfach et al. [21] to PLCs through 1,512,877 edges. By analysing the main centrality parameters computed from this network, we identified 1022 PLCs (Supplementary Table 3) that were highly connected to MAGs (with high degree, betweenness and strength) and referred thereafter as generalist or keystone PLCs (keyPLCs) (Supplementary Fig. 10). Overall, these keyPLCs had specific features; 9% of these PLCs harboured CAS proteins compared to 0.51% for the whole metaplasmidome and 78 PLCs among these key-PLCs targeted 44,087 PLCs (8.7% of the PLCs). Finally, 142 keyPLCs (13.8%) harboured CRISPR elements. In addition, 405 (39.6%) contained ARGs, and 435 (42.5%) were linked to a MAG classified by its taxonomy as a putative human pathogen. These two features represented only 7.9% and 4.1% of the total network, respectively. These keyPLCs were thus enriched in ARGs and potentially involved in pathogenicity and human health. Two keyPLCs, unknown in the reference databases, are presented as examples in the supplementary materials (Supplementary Figs. 11 and 12).

In order to quantitatively identify the main pathways of ARG dispersal, reads were recruited to keyPLCs carrying ARGs and their associated MAGs within the network. As expected, a significant correlation was observed between the recruited reads (r=0.6, P<0.05). To focus on the main ecosystems involved of the dissemination of ARGs, the mapping was restricted to the nodes of this network involved in the human health (Fig. 3). Overall, these results highlight the important role of air and wastewater in the dissemination of ARGs. Other biomes, such as freshwater and animals, may also contribute to the spread of these plasmids, albeit to a lesser extent. Surprisingly,

(See figure on next page.)

Fig. 3 Main ecosystems involved in the spread of ARGs with an impact in the human health inferred from host of plasmids (MAGs) (**A**) and keyPLCs (**B**). These Sankey diagrams were constructed from recruited reads on MAGs and keyPLCs that met the following criteria: keyPLCs with ARGs, MAGs linked to these keyPLCs with a taxonomic affiliation to a putative pathogen, and nodes detected in human ecosystems (keyPLCs or MAGs detected in human ecosystems inferred from metadata). The width of the link was directly related to the base covered by the mapping, normalised by the number of the reads mapping for each ecosystem





terrestrial and urban environments do not seem to have a strong connection to human health, whereas the ice ecosystem was involved.

#### The distribution of the PLCs across the world

The MAGs, potentially involved in human health, are distributed globally, spanning across low to high-income countries (Supplementary Fig. 13). The geographical localisation of MAGs [21] in the original publication and PLCs in this study allowed to compute the geographical distances associated with the network edges. As a result, the distances between MAGs and PLCs varied from 0 to 19,964 km, with an average distance of 7770 km, connecting different geographical locations across the planet. A similar pattern emerges when looking specifically at the most important PLCs related to human health (Fig. 4).

Finally, this study highlights few ubiquitous PLCs (i.e. 360 present in at least 8 biomes), the majority being restricted to a specific biome, and have a large global distribution. Plasmid-hosting bacteria may therefore belong to generalist taxa that are able to thrive in different environments and thus play a key role in the widespread spread of ARGs. To test this hypothesis, metagenomic reads were aligned against microbial species (MAGs clustered with an ANI > 0.95), included in the network. These species were typically found in a median of two habitats, although there were cases where they were present in up to 18 different habitats. Interestingly, microbial species interacting with keyPLCs had a significantly broader distribution across ecosystems (P < 0.001) and

biomes (P < 0.001) when compared to other microorganisms within the network. Among these different biomes, it is particularly interesting to examine the relationships with the human biome. The taxonomic composition of bacteria found in human metagenomes and at least in four other biomes was predominantly characterised by Proteobacteria, Firmicutes, Bacteroidetes and Actinobacteria. In particular, certain bacteria belonged to known pathogen species such as Pseudomonas putida, Klebsiella pneumoniae, Acinetobacter baumannii, Campylobacter jejuni, Bacteroides fragilis or Alcaligenes faecalis (Supplementary Fig. 14). Furthermore, the median ANI distance between each MAG associated with each keyPLC was 0.10 (0-0.28), and the taxonomy inferred from these MAGs revealed that 474 keyPLCs were associated with microbial species belonging to at least 2 phyla.

## Horizontal gene transfer within the metaplasmidome and with microorganisms

The PLCs in the network were randomly sampled and the recent horizontal gene transfer (HGT with 100% identity between genes) with bacterial chromosomes (i.e. MAGs) linked to these PLCs was evaluated. On average, this transfer represented 4.35% ( $\pm$ 0.68%) of the predicted genes of these PLCs from a random sampling. Considering only the keyPLCs in the network, this proportion was significantly higher, reaching 16.1% and involving 4820 MAGs. The ARG represented 9.2% of the recent transfers between MAGs and PLCs. Among the top 10 most represented genes (excluding the category 'hypothetical



**Fig. 4** Relationships inferred from microorganisms (MAGs) and keyPLCs involved in the human health. The red lines link MAGs and plasmids from the geographical localisation of MAGs in the original publication [21] and keyPLCs in this study. The names of biomes were adjusted to align the nomenclature of this paper with that of Nayfach et al. [21]

protein'), the ARGs were represented by three antibiotic efflux pumps and two genes conferring resistance to gly-copeptide antibiotics (VanS and VanR) (Supplementary Table 3).

Recent HGT between PLCs involved 240,846 transfers greater than 500 bp and 137,511 greater than 1000 bp with 100% identity. By dereplicating the HGT involved fragment greater than 1000 bp, 66,076 DNA fragments were analysed in depth. Firstly, 2.9% of these PLCs contained specific mobile genetic elements, mainly insertion sequences (IS) and inverted repeats (IR), and were therefore enriched in these elements. According to the metaplasmidome annotation, transposase (KO7497) was the most represented gene in this transfer. These horizontal gene transfers did not preferentially involve antibiotic resistance genes (ARGs), although these genes accounted for 1.8% of all genes, compared to 2.44% for the entire plasmidome. Among the resistance mechanisms, ABC transporter and glycopeptide resistance dominated, following the general pattern found in the entire metaplasmidome.

#### **Discussion and conclusions**

This study marks the pioneering effort to unravel the plasmid landscape through metagenomic experimentation, operating on a global scale spanning diverse habitats unified by biomes. Historically, plasmid investigations have predominantly focused on host cultivation, particularly in the context of pathogenic forms that impact human, animal and plant health. Other studies have used specialised protocols to isolate plasmid fractions from chromosomal ones prior to sequencing (cited in Hilpert et al. [19]). However, the available data from such approaches are limited, and the experimental protocols are not yet standardised. Furthermore, plasmids often remain ungrouped during binning due to their high copy numbers and distinct genomic signatures from their hosts, and are rarely included in MAGs [24]. Few of the PLCs identified in this study have been previously referenced in public databases or included in MAGs as defined in the study by Nayfach et al. [21], whereas it has been estimated that almost 50% of bacteria carry more than one plasmid [25]. Overall, the comparative analysis performed in the current study shows a weak similarity between data newly acquired and those from public databases and recent metaplasmidome surveys, except for data related to human ecosystems. Finally by using existing or new bioinformatics tools [26, 27] or by combining different approaches (i.e. this study), these new data-driven studies allow to better understand the characteristics of plasmids in the environment (gene content, ecology...) and to expand dedicated databases [28]. However, detecting plasmids in metagenomes presents several challenges. Plasmid sizes can vary widely, ranging from approximately 1 kbp to over 1 Mb, and assemblies are often fragmented [19]. As a result, these extrachomosomal elements are frequently incomplete and classified either based on the presence of hallmark genes, when identifiable, or solely on sequence signatures such as k-mer patterns [29]. Therefore, the presence of contaminants in the PLC database built in this study cannot be ruled out. Moreover, intriguing mobile elements such as phage-plasmids [30] may have been excluded from this analysis. For the human biome, the findings align with previous research, suggesting that the proportion of plasmids in metagenomes is around 20% such as in the gut or oral cavity [31]. The majority of proteins encoded by PLCs were of unknown features. The metaplasmidome's gene content included accessory genes also present in chromosomes but differed significantly, as expected, in the genes involved in conjugation. CAS proteins are likely to be involved in the war between MGE [32]. Thus, the CAS proteins were detected in 3% of plasmids referenced in the RefSeq database. In this metaplasmidome, the percentage was lower; however, the class 1 system was overwhelmingly predominant according to this previous study. The enrichment of keyPLCs in CAS-CRISPR could therefore explain their ecological success.

#### ARGs harboured by the metaplasmidome

The presence of ARGs in plasmids is not extensively documented in the literature and can vary depending on the annotation methodology. The ARGs can indeed have a low identity with public databases, demonstrating that the vast majority of the functional resistance genes in ecosystems represent novel sequences [33]. From different resistome studies performed on metagenomic data (i.e. resistome), Nesme et al. [34] reported a range of abundances from 0.05% in chicken gut to 5.6% in human faeces. More specifically, according to this study, Yu et al. [27], by exploring human metaplasmidome, found numerous genes encoding antibiotic resistances including mostly efflux pumps and genes targeting specific classes of antibiotics, such as glycopeptides. The abundance of antimicrobial resistance, which is linked to the selection and persistence of plasmids in an environment, may be dependent on the antibiotic concentrations. Based on the limited data available in the literature, the quantification of ARG (i.e. RPKM) in the metaplasmidome can be related to the antibiotic load in the aquatic ecosystem as this concentration increases from the sea, to rivers and polluted aquatic ecosystems as WWTP. Intriguingly, our study revealed a high abundance of ARGs in environments, such as the insect gut metagenome, where the antibiotic load is thought to be low. Nevertheless, insects harbour ARGs, including efflux pumps

which confer resistance to antibiotics when transferred to Escherichia coli [7]. Similarly, animals such as rats and mice are characterised by high levels of ARGs. Some studies suggest that plasmid-mediated resistance may be favoured in environments with relatively low antibiotic concentrations [4]. These results suggest that antibiotic concentration may not be the primary explanatory factor for ARG abundance in this context. Antibiotic concentrations in the environment are relatively low (e.g. ranging from 10 ng/L to 10  $\mu$ g/L in aquatic environments) [35], which are significantly lower than therapeutic doses in human blood plasma. Other factors, such as the linkage of ARGs and metal resistance genes in plasmids, allow heavy metals to act as a co-selective agent in the emergence and spread of antibiotic resistance [36]. However, this study shows that there is no significant relationship between these two types of resistance at the metagenome scale. Instead, according to an empirical model [4], ARG abundance was mainly explained by bacterial richness (r = 0.54, P < 0.001) (supplementary material Fig. 8B). This suggests that higher microbial diversity increases the likelihood of finding a host that supports the persistence of a particular plasmid. Thus, bacterial diversity does not appear to be an obstacle to the spread of plasmid-associated antimicrobial resistance, contrary to the conclusions of Cuadrat et al. [37] based on the entire resistome.

#### The gene flow between the ecosystems seems limited

The gene content of the PLCs, including ARGs, is intricately linked to the type of ecosystem or biome, and consequently, to the taxonomic composition. These findings suggest that the dynamics of plasmids in environment are identical regardless of the presence or absence of ARGs. From an in-depth study of human gut, Yu et al. [27] evidenced rather a strong biogeographic signal at the individual level and in the lifestyle of the populations. However, by studying the gene exchange that connects the human microbiome, Smillie et al. [15] show also that the plasmid network is shaped by ecology rather than geography. This new study provides evidence for such a pattern across multiple habitats, confirming that plasmids can be found within the same ecosystem type but in different geographical locations. Similarly, using bacterial isolates from the human gut, Yang et al. [14] found that plasmids were shared between bacterial hosts across geographically distant environmental niches. This observation may be explained by increased human contact on a global scale, particularly during the era of industrialisation. However, it is important to consider that aquatic environments and air habitats also act as mobile platforms, facilitating the spread of genes such as ARGs and accessory genes across the globe.

However, certain plasmids carrying ARGs have been identified as being shared between biomes, and knowledge of the sources and reservoirs of ARGs is essential for controlling their spread. While most studies draw conclusions from the resistome, this study allows us to identify the main pathways of plasmids from different habitats to humans. The human biome was linked to wastewater and, more interestingly, to freshwater, animals, urban and air. More specifically, the keyPLCs highlighted the importance of air and wastewater as pathways for the spread of ARGs. In particular, urban and wastewater environments were logically associated with human biomes, as these biomes can share microorganisms through simple contact and faecal discharge, respectively. The strong associations of certain biomes, including human, with animals was also expected as domestic animals are affected by antibiotics treatments. Interestingly, terrestrial ecosystems showed limited PLCs sharing with other biomes, including the human biome, and did not appear to play a role in the dissemination of ARGs to humans. This is surprising given that some studies have suggested that terrestrial ecosystems harbour the most diverse pool of ARGs [34] and have been proposed as a reservoir of resistance genes available for exchange with clinical pathogens [38]. Soil itself is a stationary complex characterised by great heterogeneity and harbours numerous ARGs, independent of human antibiotic use [7]. Nevertheless, physical forces such as air can transport soil particles, including bacteria, on an intercontinental scale. The marine environment was characterised by few ARGs as expressed in RPKM, regardless of location whereas the analysis of TARA Oceans data [39] seems to highlight numerous ARGs in plasmids. Only certain zones with high anthropogenic activity had a significant impact on marine habitats [40]. This environment shared PLCs, with or without ARGs, mainly with freshwater ecosystems. Bacterial adaptation to saline or non-saline environments involves specific evolutionary processes, such as a metaproteome characterised by different isoelectric points [41]. The gene flux between saline and non-saline environments could therefore be classified as a rare event.

## Keystone PLCs travel across ecosystems as hitchhikers embedded in the BHRs

The network established between PLCs and bacterial MAGs highlights keyPLCs that are characterised by an enrichment in antibiotic resistance genes (ARGs) and CAS-CRISPR components. In addition, certain well-documented pathogens carrying ARGs have the ability to cross different environments before posing a threat to humans. Among the pathogens associated with these specific keyPLCs, bacteria belonging to the ESKAPEE

pathogens were found. Among these pathogens, *Acine-tobacter baumannii* serves as a prominent case study, having evolved from a state of complete antibiotic susceptibility to multidrug resistance [42]. These genes have been horizontally transferred from a variety of bacterial genera, with some originating from environmental sources. *Klebsiella pneumoniae* is known to have been the first isolate to carry carbapenem resistance genes, which are critical for treating serious infections and combating multidrug-resistant Gram-negative bacterial infections [43]. Other bacteria were found to be associated with keyPLCs, such as those in the genus *Kluyvera* and *Ochrobactrum*, which are known to harbour resistance genes with a high degree of genetic similarity to those found in pathogenic microorganisms [44].

There are two main explanations for the ecological success of such plasmids. First, the majority of bacteria associated with these keyPLCs belong to versatile taxa that are able to thrive in the microbiomes of different habitats, thereby facilitating communication between different microbiomes. These taxa act as 'microbial hubs' within scale-free networks, connecting a wide range of microbiota [13]. These keyPLCs were associated with bacteria that have a broad distribution across different ecosystems. As a result, these plasmids facilitate the spread of ARGs across diverse habitats worldwide, using bacteria-including many pathogens-as hosts for direct transmission. In addition, these plasmids could be transmitted indirectly through commensal bacteria inhabiting the human digestive system, eventually reaching pathogens that cause human disease. Second, these plasmids have the ability to replicate in a wide range of hosts, potentially driving gene exchange over different phylogenetic distances. Such broad host-range plasmids were detected among keyPLCs, as the genetic distance of the bacteria associated with these plasmids can span at least two phyla. Rahube et al. [45] highlighted the transfer of IncP- and IncPromA-type plasmids from Gammaproteobacteria to many different recipients belonging to 11 different bacterial phyla in terrestrial environment. BHRs associated with ARGs have been frequently isolated from habitats such as soils [46], wastewater [47] or rivers [48].

It is reasonable to assume that these plasmids confer benefits on their hosts under antimicrobial pressure. This observation, as noted by Wang and colleagues [49], is particularly evident for antimicrobial resistance plasmids in WWTPs, where they may confer a selective advantage. However, these new data show that this conclusion can be drawn independently of the presence of antibiotics, even at low concentrations. The main ARGs detected in this study were related to efflux pumps. These genes confer antibiotic resistance but are known to have a more general role in the environment, such as pumping various biocides (metals, toxins, etc.) out of the cells and conferring a positive fitness effect to the host [50]. In addition, this association, BHR-keyPLCs, certainly promotes gene transfer between distant lineages. Specifically, this study focuses on plasmid-host transfer and sheds light on a hot spot of HGT. Recent gene transfers highlighted in this study mainly involve ARGs and multidrug export. Transfer to a plasmid has several advantages for the host. For example, many ARGs may not confer resistance when immobilised on the genome, but they do when carried by plasmids [8]. Interplasmid ARG transfer could also accelerate the dissemination of the antibiotic resistance in bacterial pathogens [51]. However, in despite the extensive analysis of numerous plasmids, this study does not provide evidence to support the promotion of such genes or specific mechanisms. Such an evolutionary strategy cannot be widespread because it must occur between compatible plasmids, i.e. that can coexist in a same bacteria over several generations.

## Identifying hot spots for ARG dissemination: key areas for surveillance

It has been argued that research, surveillance and intervention strategies to mitigate the spread of resistance require a comprehensive approach that recognises the interconnectedness of human health with that of animals and our shared environment. Thus, soil microbiota has been suggested as one of the ancient evolutionary sources of antibiotic resistance and are often considered as potential reservoirs. This study suggests that this environment is not the primary pathway for ARG dissemination, particularly when focusing on the mobilome. By comprehensively analysing the plasmidome content of different ecosystems and tracing the main route of plasmid pBI143, three habitats emerged as particularly significant: wastewater, as expected, together with freshwater (mainly river) and air. Additional findings highlight the importance of freshwater and air environments for gene flow. While previous studies have highlighted the prevalence of recombination sites (*att*C) in marine ecosystems compared to human ecosystems [52], this current study reveals the highest ratio of recombination sites in the metagenomes of freshwater and air ecosystems. Freshwater ecosystems, particularly rivers, are recognised as important hubs for ARGs and play a critical role in facilitating horizontal gene transfer and the evolution of resistance [53]. These ecosystems are affected by the input of antibiotics from various sources, including human urine and faeces, and animal manure used as fertiliser. However, air as a vector for ARG dissemination is often overlooked in surveillance efforts, and its role in plasmidmediated gene flow has been largely neglected. Thus, animal faeces, hospital and WWTPs could be one of the major sources of airborne ARGs (see review Segawa et al. [54]). For example, the bioaerosol generated by municipal sewage can potentially travel many kilometres and be deposited on soil and water [55]. The airborne bacteria and birds could be responsible for the anthropogenic ARGs detected in remote glaciers [56]. This relationship between glacier and air could explain the involvement of the ice ecosystem in the spread of ARG (Fig. 3).

#### Conclusion

This study significantly expands the database of plasmid sequences, highlighting that a significant number of genes remain uncharacterised. In addition, it provides critical insights into the primary pathways for plasmidtransmitted ARGs, thereby improving our understanding of their spread across different environments. Of particular note is the potential for air to act as a vector for long-distance transport of ARGs across ecosystems and continents. It is therefore essential to extend studies to the air and to include this environment in surveillance networks to better monitor and control the spread of antibiotic resistance.

#### Methods

The methods used in this study are presented graphically in the supplementary materials (Supplementary Fig. 15).

#### Datasets and metadata

Plasmid content was predicted from assembled data already publicly available or constructed from reads for this study. The assembled data supplied by Tully and colleagues [57], metasub consortium [58] and TARA ocean [59] were used for the human microbiome, the built environment and the marine ecosystem, respectively. For assembly in the current study, reads from metagenomes were selected from two main databases. For the soil ecosystem, the metagenomes were selected from the dedicated curated database 'TerrestrialMetagenomeDB' [60]. For the other environments, metagenomes were selected from the SRA metadata. Data were manually curated to remove metabarcoding data and retrieve some GPS locations from original publications. Some metadata specific to the human gut [22] were extracted from the following repository: https://gmrepo.humangut.info/data. All the data (accession numbers and main metadata) are summarised in Supplementary Table 5 and Supplementary Fig. 16. In addition, the taxonomic composition of the metagenomes was assessed using metaphlan (v3) [61] with a subsample of 20,000,000 reads. Bacterial richness corresponded to the number of species assessed by this tool.

#### Plasmid prediction and clustering

If the metagenomes were not assembled, reads were assembled by using megahit 1.2.9 with the metalarge option [62] after cleaning the data with bbduk2 (qtrim=rl trimq=28 minlen=25 maq=20 ktrim=r k=25 mink=11 and a list of adapters to remove) from the bbtools suite (https://jgi.doe.gov/data-and-tools/ software-tools/bbtools/) (Supplementary Fig. 15A).

were predicted for each assembly Plasmids (length > 2 kb) (Supplementary Fig. 15B) by using both reference-based and reference-free approaches (Supplementary Fig. 17) as described in previous works [19, 63] and available on the github website (https://github. com/meb-team/PlasSuite/). The databases used for the first approach included those for chromosomes (archaea and bacteria) and plasmids from RefSeq, as well as the MOB-suite tool [64], SILVA [65] and phylogenetic markers hosted by chromosomes [66]. The database created for this purpose is available at this address https://github. com/meb-team/PlasSuite/?tab=readme-ov-file#1-prepa re-or-download-your-databases. Two reference-free methods were applied to contigs that were not affiliated with chromosomes (discarded) or plasmids (retained in the first step): PlasFlow [29] and PlasClass [67]. Previously undetected viruses were removed by using Viral-Verify (https://github.com/ablab/viralVerify) [68] that provides in parallel plasmid/non-plasmid classification. This step would also remove potential plasmid-phage elements as described by Pinilla-Redondo et al. [30], but would minimise false positives. Eukaryotic contamination was removed by aligning the sequences against the NT database and human chromosomes (GRCh38) using minimap2 [69] with -x asm5 option. Contigs mapping with 95% identity for at least 80% coverage were removed. The predicted plasmids, hereafter referred as plasmidlike sequences (PLSs), were grouped by 'scientific names' (i.e. 27) such as defined in the SRA metadata (air, lake, wetland...) and subsequently named ecosystems. These ecosystems were grouped in 9 biomes (Supplementary Table 5). The data were then dereplicated by ecosystems using cd-hit-est with a threshold of 99%. The dereplicated PLSs were then clustered using MMseqs2 [70] with 80% of coverage an 90% of identity (-min-seq-id 0.90 -c 0.8 cov-mode 1 -cluster-mode 2 -alignment-mode 3 -kmerper-seq-scale 0.2) to define plasmid-like clusters (PLCs).

#### **Functional annotations**

Gene annotation of PLCs (Supplementary Fig. 14C) was based on prokka [19] (https://github.com/meb-team/ PlasSuite/tree/master/PlasAnnot), using a dedicated database specifically designed for the identification of plasmid markers and ARGs using ResFAM [71]. In a subsequent step, predicted proteins were subjected to functional annotation by alignment to the PFAM database using hmmer3 (-cut\_ga) [72] and KEGG using KoFamScan (https://github.com/takaram/kofam\_scan). Only significant KEGG results as determined by this last tool were selected in the final annotation. The BacMet database (version 2.0) was used for antibacterial biocide (BRG) and metal resistance gene (MRG) predictions [73]. Comparison with this database was performed using MMseqs2 (easy-search -s 5.7 -e 1e-3 -c 0.5 -cov-mode 1) with an identity cut-off of 70% and a coverage of at least 50% [52].

Cas genes were identified using the CRISPRCasFinder program [74] within a singularity container (https:// crisprcas.i2bc.paris-saclay.fr/Home/Download). Integron (intI) prediction was processed from genes with a specific hmmer profile (intI.hmm) available at https:// git-r3lab.uni.lu/susheel.busi/intonate [75]. Gene mobility in integrons is controlled by the presence of *att*C sites, which are 55 to 141 nucleotide long imperfect inverted repeats. These sites within this large DNA dataset were identified from PLCs and reads (i.e. metagenomes) using a hidden Markov model [76] (https://github.com/ maribuon/HattCI). The outcomes were filtered based on Viterbi scores (Vscore in the output file) greater than 7.5 to minimise false positives according to the original paper. MobileElementFinder [77] was used to detect the following types of MGEs: insertion sequences (IS), unittransposons (Tn), composite transposons (ComTn), integrative mobilisable elements (IME) and miniature inverted repeats (MITEs). Finally, conjugative plasmids (PLCs) were classified using plascad (https://github.com/ pianpianyouche/plascad) [78].

Functional annotation comparisons between PLCs and microbial chromosomes were based on predicted genes derived from the reads to identify specific features. First, reads mapped to PLCs were distinguished from unmapped reads, which were considered to be chromosomal. Protein prediction was performed on the reads using FragGeneScan [79], and the proteins were annotated against PFAM and KO databases. These annotations were then compared using the DSeq2 tool [80].

#### **Backmapping and PLC coverage**

Clean reads were mapped to the PLCs using bwa [81] and the results were merged using samtools [82] to retain reads with an identity greater than 95% (msamtools filter -b -l 50 -p 95 -z 80 –rescore –besthit) (https://github.com/arumugamlab/msamtools) (Supplementary Fig. 15D). The mappings (or recruitment reads) to the predicted genes were deduced from the gff file according to the procedure described by the OSD consortium (https://github.com/MicroB3-IS/osd-analysis/wiki/

OSD-assemblies). Briefly, this pipeline uses gff2bed, bam-Tobed and bedtools [83] to affiliate reads to a specific gene. In this step 12,833 metagenomic experiments (i.e. samples) were considered (Supplementary Table 5) each using a subset of 2,000,000 reads to reduce the computational burden (e.g. Tully et al. [57]). The average coverage of PLCs for each ecosystem was calculated from these 'bam' files using samtools (v1.16.1) [82] with the coverage command. For mapping to the plasmid pBi143 identified among the PLCs, a marker of human faecal contamination [23], all reads for all metagenomes were mapped and the results were normalised by the bp total in the sample.

From this backmapping on PLCs and genes, two matrices were generated (Supplementary Fig. 15E), representing predicted  $PLCs \times metagenomes$  or genes  $\times$  metagenomes expressed in reads or base pair (bp) counts. The row of this second table can then be sorted to select genes of interest as, for example, ARGs or KOs. These matrices were filtered to remove spurious occurrences of PLCs or genes. A PLC or gene was excluded from an ecosystem if it was either present as a singleton in all samples, or present in less than 1% of the metagenomes.

#### Gene abundance calculation

From the generated gene matrix, the rows (i.e. genes) associated with genes of interest (e.g. ARGs) were selected. The columns were normalised by the gene length and the number of clean reads. The data were multiplied by 1,000,000 to obtain the RPKM. From the ARG affiliation (ARO ontology), a resistance mechanism can be inferred and displayed by a heat map using the R package.

#### **Computation of distances**

Computing the Bray–Curtis distance on such large dimensional PLC/gene×metagenome matrices requires significant computational resources (Supplementary Fig. 15F). To overcome this, the matrices were divided into smaller parts (50 columns each) and all combinations (part against part) were parallelised on a high performance computing (HPC) system. The result was a 12,833×12,833 matrix which was analysed by NMDS, ANOSIM and Mantel statistics using the vegan package [84].

### **Comparative analyses**

The PLSs identified in this study were compared with various public databases using minimap2 [69] with the parameters '-x asm5' and requiring at least 80% coverage. Bray–Curtis distances between plasmids from the PLS grouped by ecosystems and different public databases (RefSeq, IMG/PR, etc.) were calculated using Simka [85]

with default settings. Subsequently, NMDS analysis was then performed using the vegan package [84] based on these distances (Supplementary Fig. 15G).

#### A bipartite network linking microorganisms to PLCs

A bipartite network was constructed using metagenomeassembled genomes (MAGs) from different ecosystems defined by Nayfach et al. [21] and the PLCs built in this study (Supplementary Fig. 15H). Edges were defined based on the presence of PLCs (i.e. mapping) in 52,515 MAGs or spacer-protospacer pairs. The PLCs in the MAGs were detected by mapping performed with minimap2 with a minimum identity and coverage of 90% and 80% [69]. This procedure allows to identify 55,536 PLCs associated with 14,358 MAGs. Two tools were used to detect CRISPR in MAGs: MetaCRT with default parameters [86] and PILER-CR (-minarray 4 -quiet -minspacer 20 -mincons 0.97) [87]. MetaCRT results were filtered using parser\_metaCRTout.R [88]. The spacers contained in MAGs contigs>2 kb with at least 3 repeats were pooled and clustered with cd-hit-est (-g 1 -T 16 -M 20000 -c 0.9 -d 0 -s 1 -aL 1 -aS 1). A total of 451,132 unique spacers>25 bp were then identified in 12,470 MAGs. A BLASTn search was performed using these non-redundant spacers as queries against the predicted PLCs. Spacer matches with at least 95% coverage and identity then defined a protospacer on PLCs and were retained for further analysis. By combining these results, a bipartite network with 23,498 MAGs and 596,156 PLCs nodes and 1,512,877 edges was built with the package igraph under R [89]. The main network parameters computed (betweenness, degree, strength and closeness) associated to the PLCs mean coverages were analysed by a principal component analysis (package ADE4 [90]) to determine the main keystone among the PLCs in the network (Supplementary Fig. 10).

#### Evaluation of the generalist taxa in the network

The ANI distances between all 23,498 MAGs in the bipartite network were calculated using Mash [91] with default parameters. The MAGs were then grouped into clusters using the 'bClust' function with a complete linkage, a part of the 'micropan' package implemented in the R software [92]. MAGs were classified as belonging to the same species if their distance was less than 0.05 [57]. A subset of reads (2,000,000) from 12,833 metagenomic experiments were aligned to the 23,498 MAGs using bwa and the results were subsequently filtered with msamtools according to the methods described above (backmapping chapter). Species coverage was calculated for each ecosystem. This value was determined as the mean coverage across the MAGs belonging to the respective cluster.

#### Horizontal gene transfer

In this work, both types of HGT were assessed: between bacteria (MAGs) and PLCs, taking into account the connection established in the network (Supplementary Fig. 15H), and between PLCs themselves.

To assess recent transfer between MAGs and PLCs, genes on chromosomal contigs of MAGs (without predicted plasmids) were predicted using Prodigal [93]. These genes were compared to the predicted genes on PLC using BLASTn [94] and only identical genes were selected. All MAGs associated with keyPLCs were analysed for recent transfers. For comparison with gene transfers not involving keyPLC, 3 random samples of 1000 PLCs were made. To analyse the recent transfer between PLCs, all PLCs were compared by BLASTn and only DNA fragments of 500 and 1000 bp with 100% identity were selected [95].

#### Abbreviations

| ARGs    | Antibiotic resistance genes          |
|---------|--------------------------------------|
| ICE     | Integrative and conjugative elements |
| IS      | Insertion sequences                  |
| keyPLCs | Keystone PLCs                        |
| MAGs    | Metagenome-assembled genomes         |
| MGEs    | Mobile genetic elements              |
| MRGs    | Metal resistance genes               |
| PLCs    | Plasmid-like clusters                |
|         |                                      |

#### Supplementary Information

The online version contains supplementary material available at https://doi. org/10.1186/s40168-025-02062-5.

Supplementary Material 1. Supplementary Material 2. Supplementary Material 3. Supplementary Material 4.

#### Acknowledgements

We are grateful to the 'Mésocentre Clermont-Auvergne' (https://mesocentre. uca.fr/) and the plateforme AuBi of the Université Clermont Auvergne and Genouest (https://www.genouest.org/) for providing help with computing and storage resources.

#### Authors' contributions

DD analyzed the data and wrote the manuscript.

#### Funding

Non applicable.

#### Data availability

The datasets analyzed during the current study are available in the Zenodo repository (v2), https://zenodo.org/records/14713180.

#### Declarations

Ethics approval and consent to participate Non applicable.

#### **Consent for publication**

Non applicable.

#### **Competing interests**

The authors declare no competing interests.

Received: 22 October 2024 Accepted: 9 February 2025 Published online: 19 March 2025

#### References

- Smillie C, Garcillán-Barcia MP, Francia MV, Rocha EPC, de la Cruz F. Mobility of plasmids. Microbiol Mol Biol Rev. 2010;74:434–52.
- Rodríguez-Beltrán J, DelaFuente J, León-Sampedro R, MacLean RC, San Millán Á. Beyond horizontal gene transfer: the role of plasmids in bacterial evolution. Nat Rev Microbiol. 2021;19:347–59.
- Smalla K, Sobecky PA. The prevalence and diversity of mobile genetic elements in bacterial communities of different environmental habitats: insights gained from different methodological approaches. FEMS Microbiol Ecol. 2002;42:165–75.
- Castañeda-Barba S, Top EM, Stalder T. Plasmids, a molecular cornerstone of antimicrobial resistance in the One Health era. Nat Rev Microbiol. 2023. https://doi.org/10.1038/s41579-023-00926-x.
- Martínez JL, Coque TM, Baquero F. What is a resistance gene? Ranking risk in resistomes. Nat Rev Microbiol. 2015;13:116–23.
- Debroas D, Siguret C. Viruses as key reservoirs of antibiotic resistance genes in the environment. ISME J. 2019;13:2856–67.
- Allen HK, et al. Call of the wild: antibiotic resistance genes in natural environments. Nat Rev Microbiol. 2010;8:251–9.
- Nielsen TK, Browne PD, Hansen LH. Antibiotic resistance genes are differentially mobilized according to resistance mechanism. GigaScience. 2022;11:giac072.
- Siguier P, Gourbeyre E, Chandler M. Bacterial insertion sequences: their genomic impact and diversity. FEMS Microbiol Rev. 2014;38:865–91.
- Botelho J, Schulenburg H. The role of integrative and conjugative elements in antibiotic resistance evolution. Trends Microbiol. 2021;29:8–18.
- 11. Berendonk TU, et al. Tackling antibiotic resistance: the environmental framework. Nat Rev Microbiol. 2015;13:310–7.
- Gillings MR. Integrons: past, present, and future. Microbiol Mol Biol Rev. 2014;78:257–77.
- Baquero F, Coque TM, Martínez J-L, Aracil-Gisbert S, Lanza VF. Gene transmission in the one health microbiosphere and the channels of antimicrobial resistance. Front Microbiol. 2019;10:2892.
- Yang, L. et al. Global transmission of broad-host-range plasmids derived from the human gut microbiome. Nucl Acids Res gkad498 (2023) https:// doi.org/10.1093/nar/gkad498.
- Smillie CS, et al. Ecology drives a global network of gene exchange connecting the human microbiome. Nature. 2011;480:241–4.
- Sommer MOA, Munck C, Toft-Kehler RV, Andersson DI. Prediction of antibiotic resistance: time for a new preclinical paradigm? Nat Rev Microbiol. 2017;15:689–96.
- Walker A. Welcome to the plasmidome. Nat Rev Microbiol. 2012. https:// doi.org/10.1038/nrmicro2804.
- 18. Kav AB, et al. Insights into the bovine rumen plasmidome. PNAS. 2012;109:5452–7.
- Hilpert C, Bricheux G, Debroas D. Reconstruction of plasmids by shotgun sequencing from environmental DNA: which bioinformatic workflow? Brief Bioinform. 2021;22:bbaa059.
- Lai S, et al. mMGE: a database for human metagenomic extrachromosomal mobile genetic elements. Nucleic Acids Res. 2021;49:D783–91.
- Nayfach S, et al. A genomic catalog of Earth's microbiomes. Nat Biotechnol. 2021;39:499–509.
- Dai D, et al. GMrepo v2: a curated human gut microbiome database with special focus on disease markers and cross-dataset comparison. Nucleic Acids Res. 2022;50:D777–84.
- 23. Fogarty EC, et al. A cryptic plasmid is among the most numerous genetic elements in the human gut. Cell. 2024;187:1206-1222.e16.
- 24. New FN, Brito IL. What is metagenomics teaching us, and what is missed? Annu Rev Microbiol. 2020;74:annurev-micro-012520-072314.
- 25. Carroll AC, Wong A. Plasmid persistence: costs, benefits, and the plasmid paradox. Can J Microbiol. 2018;64:293–304.

- Camargo AP, et al. Identification of mobile genetic elements with geNomad. Nat Biotechnol. 2023;42:1–10. https://doi.org/10.1038/ s41587-023-01953-y.
- Yu MK, Fogarty EC, Eren AM. Diverse plasmid systems and their ecology across human gut metagenomes revealed by PlasX and MobMess. Nat Microbiol. 2024;9:830–47.
- 28. Camargo, A. P. et al. IMG/PR: a database of plasmids from genomes and metagenomes with rich annotations and metadata. Nucleic Acids Research gkad964 (2023) https://doi.org/10.1093/nar/gkad964.
- Fang Z, et al. PPR-Meta: a tool for identifying phages and plasmids from metagenomic fragments using deep learning. GigaScience. 2019;8:giz066.
- Pinilla-Redondo R, et al. Type IV CRISPR–Cas systems are highly diverse and involved in competition between plasmids. Nucleic Acids Res. 2020;48:2000–12.
- 31. Munck C, et al. Limited dissemination of the wastewater treatment plant core resistome. Nat Commun. 2015;6:8452.
- Nesme J, et al. Large-scale metagenomic-based study of antibiotic resistance in the environment. Curr Biol. 2014;24:1096–100.
- Larsson DGJ, Flach C-F. Antibiotic resistance in the environment. Nat Rev Microbiol. 2022;20:257–69.
- Baker-Austin C, Wright MS, Stepanauskas R, McArthur JV. Co-selection of antibiotic and metal resistance. Trends Microbiol. 2006;14:176–82.
- 35. Klümper, U. et al. Microbiome diversity: a barrier to the environmental spread of antimicrobial resistance? 2023.03.30.534382. 2023. Preprint at https://doi.org/10.1101/2023.03.30.534382.
- Forsberg KJ, et al. The shared antibiotic resistome of soil bacteria and human pathogens. Science. 2012;337:1107–11.
- Cuadrat RRC, Sorokina M, Andrade BG, Goris T, Dávila AMR. Global ocean resistome revealed: exploring antibiotic resistance gene abundance and distribution in TARA Oceans samples. Gigascience. 2020;9:giaa046.
- Furlan JPR, et al. Appearance of mcr-9, blaKPC, cfr and other clinically relevant antimicrobial resistance genes in recreation waters and sands from urban beaches. Brazil Mar Pollut Bull. 2021;167:112334.
- Cabello-Yeves PJ, Rodriguez-Valera F. Marine-freshwater prokaryotic transitions require extensive changes in the predicted proteome. Microbiome. 2019;7:1–12.
- Fournier P-E, et al. Comparative genomics of multidrug resistance in acinetobacter baumannii. PLoS Genet. 2006;2:e7.
- 41. Yong D, et al. Characterization of a new metallo-beta-lactamase gene, bla(NDM-1), and a novel erythromycin esterase gene carried on a unique genetic structure in Klebsiella pneumoniae sequence type 14 from India. Antimicrob Agents Chemother. 2009;53:5046–54.
- 42. Farmer JJ, et al. Kluyvera, a new (redefined) genus in the family Enterobacteriaceae: identification of Kluyvera ascor-bata sp. nov. and Kluyvera cryocrescens sp. nov. in clinical specimens. J Clin Microbiol. 1981;13:919–33.
- Klümper U, et al. Broad host range plasmids can invade an unexpectedly diverse fraction of a soil bacterial community. ISME J. 2015;9:934–45.
- 44. Heuer H, et al. IncP-1ε plasmids are important vectors of antibiotic resistance genes in agricultural systems: diversification driven by class 1 integron gene cassettes. Front Microbiol. 2012;3:2.
- Rahube TO, Viana LS, Koraimann G, Yost CK. Characterization and comparative analysis of antibiotic resistance plasmids isolated from a wastewater treatment plant. Front Microbiol. 2014;5:558.
- De la Cruz Barrón M, Merlin C, Guilloteau H, Montargès-Pelletier E, Bellanger X. Suspended materials in river waters differentially enrich class 1 integron- and IncP-1 plasmid-carrying bacteria in sediments. Front Microbiol. 2018;9:1443.
- Risely A, et al. Host-plasmid network structure in wastewater is linked to antimicrobial resistance genes. Nat Commun. 2024;15:555.
- Blanco P, et al. Bacterial multidrug efflux pumps: much more than antibiotic resistance determinants. Microorganisms. 2016;4:14.
- Wang X, et al. Inter-plasmid transfer of antibiotic resistance genes accelerates antibiotic resistance in bacterial pathogens. ISME J. 2024;18:wrad032.
- Buongermino Pereira M, et al. A comprehensive survey of integronassociated genes present in metagenomes. BMC Genomics. 2020;21:495.
- Ahmad N, Joji RM, Shahid M. Evolution and implementation of One Health to control the dissemination of antibiotic-resistant bacteria and resistance genes: a review. Front Cell Infect Microbiol. 2023;12:1065796.

- Kormos D, Lin K, Pruden A, Marr LC. Critical review of antibiotic resistance genes in the atmosphere. Environ Sci Processes Impacts. 2022;24:870–83.
- Gaviria-Figueroa A, Preisner EC, Hoque S, Feigley CE, Norman RS. Emission and dispersal of antibiotic resistance genes through bioaerosols generated during the treatment of municipal sewage. Sci Total Environ. 2019;686:402–12.
- 54. Segawa T, et al. Distribution of antibiotic resistance genes in glacier environments. Environ Microbiol Rep. 2013;5:127–34.
- Pasolli E, et al. Extensive unexplored human microbiome diversity revealed by over 150,000 genomes from metagenomes spanning age, geography, and lifestyle. Cell. 2019;176:649-662.e20.
- Danko, D. et al. Global genetic cartography of urban metagenomes and anti-microbial resistance. bioRxiv 724526 (2020) https://doi.org/10.1101/ 724526.
- Tully BJ, Graham ED, Heidelberg JF. The reconstruction of 2,631 draft metagenome-assembled genomes from the global oceans. Scientific Data. 2018;5:170203.
- Corrêa FB, Saraiva JP, Stadler PF, da Rocha UN. TerrestrialMetagenomeDB: a public repository of curated and standardized metadata for terrestrial metagenomes. Nucleic Acids Res. 2020;48:D626–32.
- Beghini F, et al. Integrating taxonomic, functional, and strain-level profiling of diverse microbial communities with bioBakery 3. eLife. 2021;10:e65088.
- Li D, Liu C-M, Luo R, Sadakane K, Lam T-W. MEGAHIT: an ultra-fast singlenode solution for large and complex metagenomics assembly via succinct de Bruijn graph. Bioinformatics. 2015;31:1674–6.
- Hennequin C, Forestier C, Traore O, Debroas D, Bricheux G. Plasmidome analysis of a hospital effluent biofilm: status of antibiotic resistance. Plasmid. 2022;122:102638.
- 62. Robertson J, Nash JHE. MOB-suite: software tools for clustering, reconstruction and typing of plasmids from draft assemblies. Microbial Gen. 2018;4:e000206.
- Quast C, et al. The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Res. 2013;41:D590–6.
- 64. Wu D, Jospin G, Eisen JA. Systematic identification of gene families for use as "markers" for phylogenetic and phylogeny-driven ecological studies of bacteria and archaea and their major subgroups. PLoS One. 2013;8:e77033.
- Krawczyk PS, Lipinski L, Dziembowski A. PlasFlow: predicting plasmid sequences in metagenomic data using genome signatures. Nucleic Acids Res. 2018;46:e35.
- Pellow D, Mizrahi I, Shamir R. PlasClass improves plasmid sequence classification. PLoS Comput Biol. 2020;16:e1007781.
- Antipov D, Raiko M, Lapidus A, Pevzner PA. MetaviralSPAdes: assembly of viruses from metagenomic data. Bioinformatics. 2020;36:4126–9.
- Pfeifer E, Moura de Sousa JA, Touchon M, Rocha EPC. Bacteria have numerous distinctive groups of phage–plasmids with conserved phage and variable plasmid gene repertoires. Nucleic Acids Res. 2021;49:2655–73.
- Li H. Minimap2: pairwise alignment for nucleotide sequences. Bioinformatics. 2018;34:3094–100.
- Steinegger M, Söding J. MMseqs2 enables sensitive protein sequence searching for the analysis of massive data sets. Nat Biotechnol. 2017. https://doi.org/10.1038/nbt.3988.
- Gibson MK, Forsberg KJ, Dantas G. Improved annotation of antibiotic resistance determinants reveals microbial resistomes cluster by ecology. ISME J. 2015;9:207–16.
- 72. Eddy SR. Profile hidden Markov models. Bioinformatics. 1998;14:755–63.
- Pal C, Bengtsson-Palme J, Rensing C, Kristiansson E, Larsson DGJ. BacMet: antibacterial biocide and metal resistance genes database. Nucleic Acids Res. 2014;42:D737–43.
- Couvin D, et al. CRISPRCasFinder, an update of CRISRFinder, includes a portable version, enhanced performance and integrates search for Cas proteins. Nucleic Acids Res. 2018;46:W246–51.
- de Nies L, et al. Evolution of the murine gut resistome following broadspectrum antibiotic treatment. Nat Commun. 2022;13:2296.
- Pereira MB, Wallroth M, Kristiansson E, Axelson-Fisk M. HattCI: fast and accurate attC site identification using hidden Markov models. J Comput Biol. 2016;23:891–902.

- Johansson MHK, et al. Detection of mobile genetic elements associated with antibiotic resistance in Salmonella enterica using a newly developed web tool: MobileElementFinder. J Antimicrob Chemother. 2021;76:101–9.
- 78. Che Y, et al. Conjugative plasmids interact with insertion sequences to shape the horizontal transfer of antimicrobial resistance genes. Proc Natl Acad Sci. 2021;118:e2008731118.
- Rho M, Tang H, Ye Y. FragGeneScan: predicting genes in short and errorprone reads. Nucl Acids Res. 2010;38:e191–e191.
- Love MI, Huber W, Anders S. Moderated estimation of fold change and dispersion for RNA-sea data with DESeq2. Genome Biol. 2014;15:550.
- Li H, Durbin R. Fast and accurate short read alignment with Burrows-Wheeler transform. Bioinformatics. 2009;25:1754–60.
- Quinlan AR, Hall IM. BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics. 2010;26:841–2.
- Li H, et al. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009;25:2078–9.
- Dixon P. VEGAN, a package of R functions for community ecology. J Veg Sci. 2003;14:927–30.
- Benoit G, et al. SimkaMin: fast and resource frugal de novo comparative metagenomics. Bioinformatics. 2020;36:1275–6.
- Rho M, Wu Y-W, Tang H, Doak TG, Ye Y. Diverse CRISPRs evolving in human microbiomes. PLoS Genet. 2012;8:e1002441.
- Edgar RC. PILER-CR: fast and accurate identification of CRISPR repeats. BMC Bioinformatics. 2007;8:18.
- Martínez Arbas S, et al. Roles of bacteriophages, plasmids and CRISPR immunity in microbial community dynamics revealed using time-series integrated meta-omics. Nat Microbiol. 2021;6:123–35.
- Csardi, G. & Nepusz, T. The igraph software package for complex network research. 1695 (2006).
- Thioulouse, J. et al. Multivariate analysis of ecological data with Ade4. (Springer New York, New York, NY, 2018). https://doi.org/10.1007/ 978-1-4939-8850-1.
- 91. Ondov BD, et al. Mash: fast genome and metagenome distance estimation using MinHash. Genome Biol. 2016;17:132.
- 92. Snipen L, Liland KH. micropan: an R-package for microbial pan-genomics. BMC Bioinform. 2015;16:79.
- Hyatt D, et al. Prodigal: prokaryotic gene recognition and translation initiation site identification. BMC Bioinform. 2010;11:119.
- 94. Camacho C, et al. BLAST+: architecture and applications. BMC Bioinform. 2009;10:421.
- 95. Groussin M, et al. Elevated rates of horizontal gene transfer in the industrialized human microbiome. Cell. 2021;184:2053-2067.e18.

### **Publisher's Note**

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.