

# Resistome analysis of bloodstream infection bacterial genomes reveals a specific set of proteins involved in antibiotic resistance and drug efflux

Willian K. Oliveira<sup>1</sup>, Mariana Ferrarini<sup>2</sup>, Luis G. Morello<sup>1</sup> and Helisson Faoro<sup>1,3,\*</sup>

<sup>1</sup>Laboratory for Applied Science and Technology in Health, Carlos Chagas Institute, FIOCRUZ, Paraná, 81350-010, Brazil, <sup>2</sup>Laboratoire de Biométrie et Biologie Évolutive, UMR 5558, CNRS, Université de Lyon, Université Lyon 1, Villeurbanne 69622, France and <sup>3</sup>Graduate Program on Bioinformatics, Federal University of Paraná, Paraná, 81520-260, Brazil

Received December 17, 2019; Revised June 30, 2020; Editorial Decision July 08, 2020; Accepted July 13, 2020

## ABSTRACT

**Bacterial resistance to antibiotics is a global public health problem. Its association with bloodstream infections is even more severe and may easily evolve to sepsis. To improve our response to these bacteria, it is essential to gather thorough knowledge on the main pathogens along with the main mechanisms of resistance they carry. In this paper, we performed a large meta-analysis of 3872 bacterial genomes isolated from blood samples, from which we identified 71 745 antibiotic resistance genes (ARGs). Taxonomic analysis showed that Proteobacteria and Firmicutes phyla, and the species *Klebsiella pneumoniae* and *Staphylococcus aureus* were the most represented. Comparison of ARGs with the Resfams database showed that the main mechanism of antibiotic resistance is mediated by efflux pumps. Clustering analysis between resistome of blood and soil-isolated bacteria showed that there is low identity between transport and efflux proteins between bacteria from these environments. Furthermore, a correlation analysis among all features showed that *K. pneumoniae* and *S. aureus* formed two well-defined clusters related to the resistance mechanisms, proteins and antibiotics. A retrospective analysis has shown that the average number of ARGs per genome has gradually increased. The results demonstrate the importance of comprehensive studies to understand the antibiotic resistance phenomenon.**

## INTRODUCTION

Antibiotics revolutionized medicine in the 1930s and 1940s, saving millions of lives. However, their indiscriminate use quickly led to the emergence of the first strains of penicillin-

resistant bacteria in 1950 (1) and, over the subsequent 40 years, to the emergence of bacteria resistant to all available antibiotics, becoming a worldwide problem (2). In general, bacterial resistance arises when a change in the genome, which may result from a spontaneous mutation and/or gene acquisition, causes biochemical changes that modify the intracellular antimicrobial concentration or the affinity of the target for the antimicrobial. These changes in the genome create the genetic variability upon which natural selection acts, giving advantages to the fittest organisms. Drugs act as selective agents, favoring the rare resistant bacteria present in the population (3). There are different classes of antibiotic resistance mechanisms (4); however, they can be divided into three main categories: (i) antibiotic inactivation-based mechanisms; (ii) antibiotic efflux-based mechanisms; and (iii) target-based mechanisms.

Antibiotic-inactivating or antibiotic-modifying enzymes are considered the main resistance mechanisms and have already been described for most antibiotics. Antibiotic inactivation can occur by transferring chemical groups to the antibiotic molecule, which reduces its affinity for the target, or by degradation involving hydrolases. In the latter case, beta-lactamases are the most studied enzymes, mainly because they are the main mechanism of resistance to beta-lactam antibiotics, the most widely used class of antimicrobials in clinical practice (5).

Changes in the expression of efflux pumps, which are membrane proteins that export antibiotics out of the cell and keep intracellular antibiotic concentrations low, are responsible for the increased clearance of antimicrobials (6,7). In gram-positive bacteria, efflux pumps consist of a single component that transfers its substrate across the cytoplasmic membrane. Multicomponent efflux pumps have been found in gram-negative bacteria; these transporters contain a periplasmic membrane fusion protein and an outer membrane protein responsible for the transfer of substrates through the cell wall. Antibiotics of all classes, except polymyxins, are susceptible to efflux sys-

\*To whom correspondence should be addressed. Tel: +55 41 3316 3230; Fax: +55 41 3316 3267; Email: helisson.faoro@fiocruz.br; hfaoro@gmail.com

tems. Efflux pumps may be drug- or class-specific and are capable of causing resistance to several different compounds (macrolides, tetracyclines and fluoroquinolones). Multidrug resistance efflux pumps are categorized into five super families based on amino acid sequence identity, the energy source used to export substances and the substrate specificities of the different pumps. These pumps include adenosine triphosphate-binding cassette (ABC), major facilitator superfamily (MFS), resistance-nodulation-division family (RND), small multidrug resistance (SMR) and the superfamily multidrug and toxic compounds extrusion (8). The target-based mechanisms occur due to mutations that lead to a reduction in the affinity of the drug; for example, quinolone and fluoroquinolone resistance due to modifications in topoisomerase II (*gyrA* and *gyrB*) and IV (*parC* and *parE*) (9,10).

Resistant microorganism infections may be responsible for common community infections such as respiratory infections and diarrhea, but resistant microorganism infections are more worrying in the hospital environment. The combination of susceptible patients with prolonged and intensive use of antimicrobials has resulted in highly resistant nosocomial infections. Resistant hospital-acquired infections are extremely difficult to treat and costly to eradicate or even control. The consequences of this scenario are severe, leading to higher morbidity, mortality and length of stay in the hospital, as well as a direct increase in hospital costs (11).

In particular, the presence of multidrug resistant bacteria in intensive care units (ICUs) not only is a frequently occurring situation but also is the most worrying situation, especially due to the high-selective pressure of antibiotics in these environments. Bloodstream infections caused by resistant gram-negative organisms can reach a mortality rate of 80 to 85%. In the EPIC II study, which included over 13 000 patients in 1200 ICUs from 75 countries, the overall prevalence of infections in ICUs was analyzed. These authors found that 51% of patients admitted to these units were diagnosed with infection and 71% were receiving antimicrobials, with the majority using two or more antibiotics (12). Currently, the rates of infections caused by resistant bacteria have been steadily increasing. In the United States, at least 23 000 deaths from multidrug resistant bacterial infections are reported each year, and the number of people affected by these infections annually is estimated at 2 million (13,14). Unfortunately, forecasts suggest that the number of deaths from these infections will grow even higher, reaching ~10 million deaths by 2050, with a cumulative cost of US \$100 trillion (14).

In this work, we performed a meta-analysis of thousands of bacterial genomes isolated from the blood of patients to identify the main mechanisms of antibiotic resistance in this state.

## METHODOLOGY

### Dataset for analysis

The organisms for the analysis were selected with the Pathosystems Resource Integration Center v3.5.34 (PATRIC) database (15). We selected bacteria that were

isolated from humans (Host name = *Homo sapiens*) and had the word 'blood' in their database records. From the total genomes deposited in the PATRIC database (227 568 genomes), this search returned a record table containing 7586 entries distributed in 6733 whole genome sequencing (WGS), 436 complete genomes and 394 plasmid sequences. This table was then filtered by python scripts so that only sequences that had blood isolation information in at least one of the following fields remained: 'Isolation site', 'Isolation source', 'Body sample site' or 'Body sample subsite'. After this filtering, the results were manually reviewed to ensure that no false positive sequences were obtained in the analysis. Sequences that did not have 'NCBI taxon id' and 'Genome Status', and plasmids that did not link to any genome or WGS sequence through the same 'Assembly accession' were filtered out, resulting in a total of 4028 final sequences (3616 WGS, 256 complete genomes and 156 plasmids). Finally, the proteomes of these genomes and plasmids were downloaded from the PATRIC database in fasta format. Due to a lack of standardization in the date format accepted by the database in the 'Collection Date' field, the data in this field were manually standardized to represent the correct date. As a control group for our analysis, we used 877 genomes of soil bacteria obtained from the RefSoil database (16). The genomes of soil bacteria were analyzed using the same methodology.

### Taxonomic distribution

To estimate the taxonomic diversity of the sample, the 'NCBI taxon id' field of the WGS sequences and complete genomes was used, thus excluding the plasmid sequences. To this end, we used a python script, which searched through the Taxonomy database HTML-NCBI (17). Taxonomic information regarding the phylum, class, order, family, genus and species levels for each of the sequences was recovered.

### Identification of antibiotic resistance genes (ARGs)

The identification of antibiotic resistance genes (ARGs) was performed using the hmmsearch algorithm (18) with the -cut-ga option against the hmm Resfams core v1.2 model (19), which has only protein data with experimentally validated antibiotic resistance functions. The results were filtered, and only the ARGs with hits that had an *e*-value < 10e-5 were retained. Nonhomologous proteins that had sufficient *e*-values to pass the cut off were removed whenever the bias value (bv) was in the same magnitude order as the sequence bit score (bs) ( $bv > bs / 10$ ). In cases where two or more hits were obtained for the same gene, only the hit with the lowest *e*-value was considered. The families of resistance genes provided by Resfams were classified using the information available in the Comprehensive Antibiotic Resistance Database (CARD) (20) according to their resistance mechanisms, type of proteins involved and classes of antibiotics to which they confer resistance. Through this classification, we categorized the genes found in our analysis and calculated the occurrence of genes for each category.

## Cluster analysis of ARGs

To verify the identity between proteins of blood and soil isolates, we performed a cluster analysis using the cd-hit program (21) with 90% identity. A composition cluster analysis was performed normalizing the number of proteins from soil through the multiplication of its occurrences by the ratio of total number of blood environment proteins by the total number of soil environment proteins in the cluster. Then, the normalized number of proteins in both conditions of each cluster was then compared to verify the proximity between proteins found in blood and soil environment and its diversity.

## Statistical analysis

Tests for independence between two features were performed based on the Chi-square test or Fisher's exact test for independence and/or homogeneity between the categorical variables in the contingency tables. In all cases,  $P$ -values  $< 0.05$  were considered statistically significant. All tests were performed in R. The packages gplots (22) and vcd (23) were used to better visualize the contingency tables for each figure.

## RESULTS

### Diversity of isolated blood bacteria

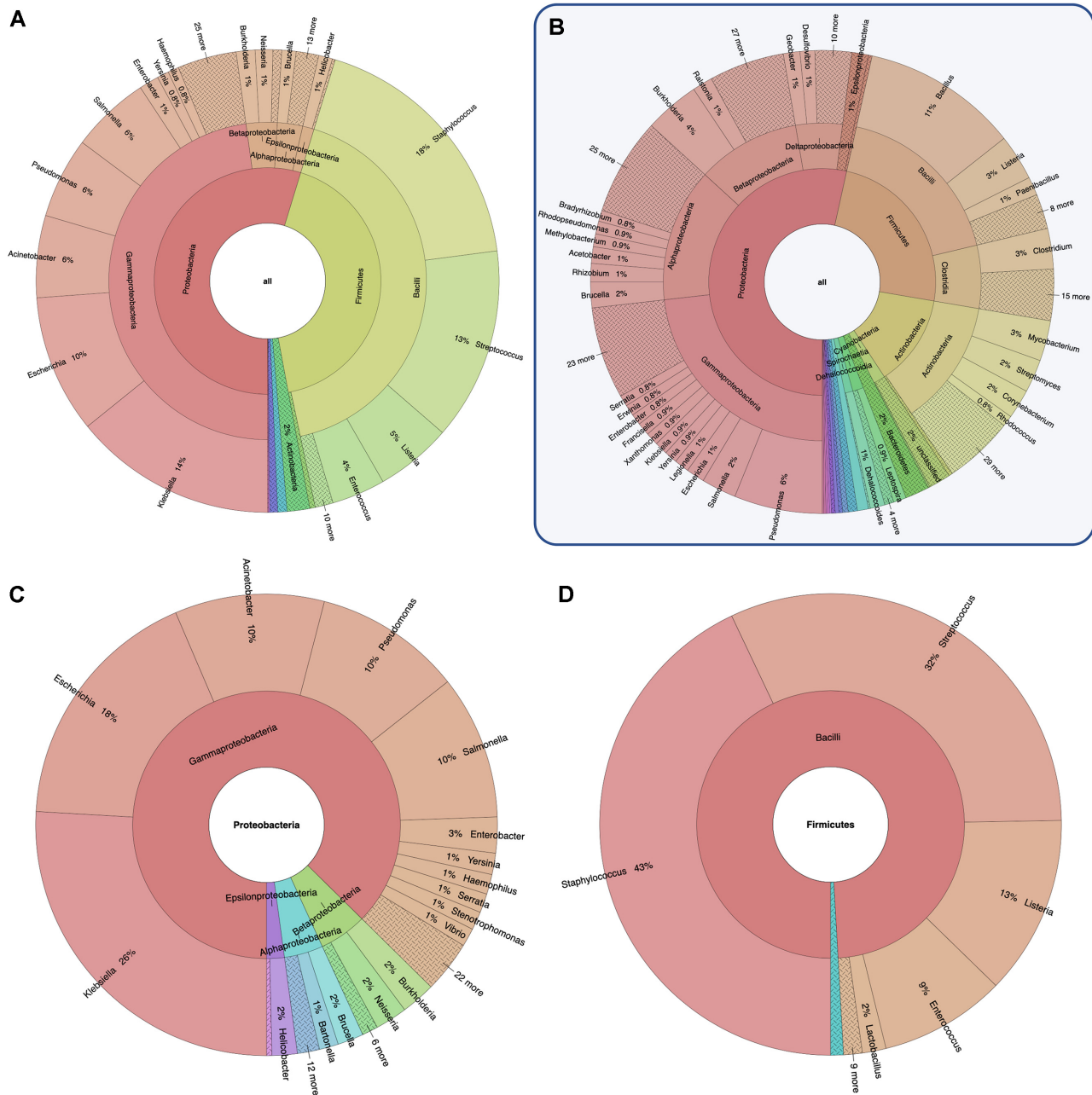
Mining of the PATRIC database for blood-isolated bacterial genomes resulted in the identification of 3872 genomes, including complete genomes and WGS. This number represents approximately 1.6% of the total genomes deposited in PATRIC, which had 227 568 genome sequences in the version used. Diversity analysis of the selected genomes revealed that they were distributed into 7 phyla, 14 classes, 105 genera and 214 species. Of this total, 56.2% of the genomes belonged to the Proteobacteria phylum, followed by the Firmicutes phylum, with 41.2% (Figure 1A). The other phyla identified (Actinobacteria, Bacteroidetes, Spirochaetes, Tenericutes and Fusobacteria) represented between 1.3 and 0.05% of the genomes. Among the most represented bacterial classes, two taxa (Gammaproteobacteria and Bacilli) accounted for 90% of the blood isolate genomes with 49 and 41%, respectively (Figure 1C and D). At the genus level, *Staphylococcus* (Firmicutes) appeared to be the most represented among blood isolate genomes, with 18.2% of the total genomes, followed by *Klebsiella* (Proteobacteria), with 15.2% (Figure 1C and D). Of the 105 genera identified, the 4 most represented contributed ~57% of the total; with two genera of gram-positive bacteria, *Staphylococcus* and *Streptococcus*, and with two genera of gram-negative bacteria, *Klebsiella* and *Escherichia*. The distribution between the different genera of the Gammaproteobacteria class in the blood isolates was more homogeneous than within the Firmicutes phylum; 44.1% of the Firmicutes genomes belonged to the genus *Staphylococcus*. Of the 214 bacterial species identified, the 8 most abundant represented more than 80% of the identified and analyzed genomes: *Staphylococcus aureus*, *Klebsiella pneumoniae*, *Escherichia coli*, *Streptococcus pneumoniae*, *Listeria monocytogenes*, *Pseudomonas aeruginosa*, *Acinetobacter baumannii*

and *Salmonella enterica*. Among these species, *S. aureus* was the most identified with 603 genomes, while *K. pneumoniae* had 536 genomes (Supplementary Figure S1).

The genomes of soil bacteria were distributed in 22 phyla, 38 classes, 282 genera and 470 species, showing a higher diversity compared to blood isolates (Figure 1B). The most representative phyla in the soil database were also Proteobacteria (53%) and Firmicutes (24%) followed by Actinobacteria (13%), Cyanobacteria (2%), Bacteroidetes (2%) and Spirochaetia (2%). Unlike the distribution of taxa found among blood isolates, there is no dominance of a specific group. While Gammaproteobacteria represented 87% of Proteobacteria among blood isolates, among soil isolates this representativeness dropped to 43%. The other classes, Alpha, Beta and Deltaproteobacteria, also showed a homogeneous distribution. The Bacilli class also showed a decrease among blood and soil isolates, going from 99 to 74% of representativeness of the phylum Firmicutes, respectively. However, the most drastic change has occurred at the genus level. The representativeness of the genera *Klebsiella* and *Staphylococcus* decreased from 15.2 and 18.2% of the total genomes of blood isolates to 0.9 and 0.6% among genomes of soil isolates, respectively. The genera *Pseudomonas* (6%) and *Bacillus* (11%) were the most represented in soil genomes. These data suggest that there is a trend for some specific groups in blood samples and indicates potential emerging isolates.

### Main mechanisms of antibiotic resistance identified in blood isolates

The identification of ARGs present in blood-isolated bacterial genomes was performed by comparison with the HMM model obtained from Resfams. To this end, the amino acid sequences of all encoded proteins were retrieved from the original PATRIC files. From the 3872 genomes, a total of 15 796 644 protein sequences were obtained. Of this total, 71 745 sequences (0.45%) were matched against Resfams in accordance with the cut off. These sequences were distributed among 10 major antibiotic resistance mechanisms: antibiotic inactivation, ABC transporter, MFS efflux, SMR efflux, RND efflux, other antibiotic efflux, target protection, target replacement, target overexpression and target alteration. However, these 10 mechanisms can be grouped into three broader categories: (i) antibiotic inactivation-based mechanism; (ii) antibiotic efflux-based mechanisms; and (iii) target-based mechanisms. Considering this distribution, the vast majority of the identified genes were within the antibiotic efflux-based category with 72.07% of the sequences, while antibiotic inactivation-based and target-based mechanisms made up 19.57 and 3.36%, respectively. A similar distribution was found when considering the amount of ARGs normalized by the number of analyzed genomes: 72.06% efflux-based, 19.57% inactivation-based and 8.36% target-based mechanisms. In our method, no genes related to the target overexpression mechanism were found. Among the mechanisms identified in blood-isolated bacterial genomes, those with the most genes assigned were RND efflux (33.74%), ABC transporter (25.18%) and antibiotic inactivation (19.57%) (Figure 2A). The same methodology was applied to the soil-isolate bacte-

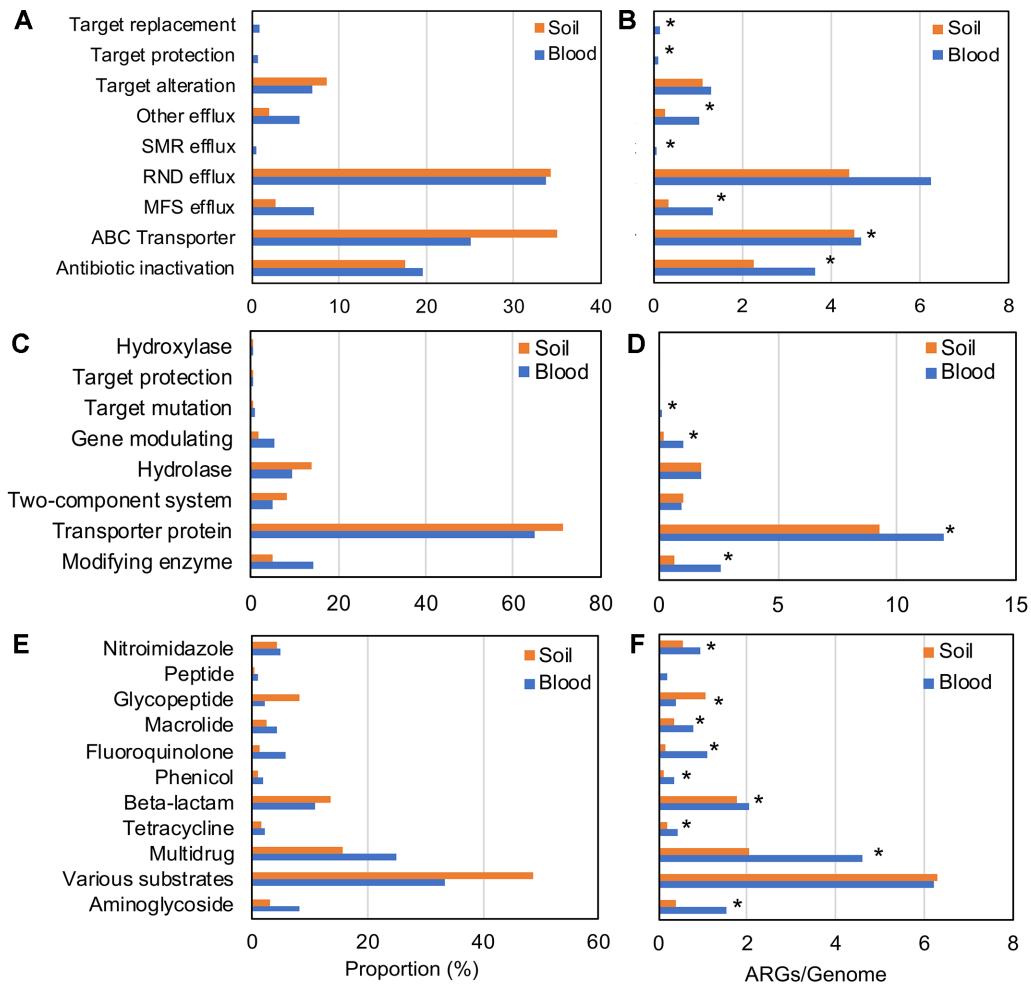


**Figure 1.** Genome diversity of blood isolates. Taxonomic classification was obtained from the NCBI TaxonID. (A) Phylum-level distribution of blood-isolated bacterial genomes. (B) Phylum-level distribution of soil-isolated bacterial genomes. (C) Taxonomic distribution within the phylum Proteobacteria of blood isolate genomes. (D) Taxonomic distribution within the Firmicutes phylum of blood isolate genomes.

rial genomes. Of the 877 genomes, 3 633 861 proteins were extracted and compared with Resfams. A total of 11 325 (0.31%) found a match in the Resfams database. The distribution of ARGs within the established classes showed that there are similarities and differences between the values obtained for blood and soil bacterial genomes. The proportion of ARGs identified as RND efflux was similar to that found for the soil (34.17%). On the other hand, the proportion of ARGs identified as ABC transporter is higher in soil-isolated bacterial genomes (34.99%). The proportion of other antibiotic efflux proteins and inactivating antibiotic

proteins is higher in the genomes of blood isolates (1.90 and 17.56%) (Figure 2A). However, when we consider the normalized value of ARGs by genome, we find that all classes of resistance mechanisms are higher in blood-isolated bacteria. With the exception of antibiotic inactivation and RND efflux, all other classes had a  $P$ -value  $< 0.05$  in  $t$ -test for two samples (Figure 2B).

The comparison between the mechanisms identified in the two most represented bacterial phyla, Proteobacteria and Firmicutes, showed a discrepancy between gram-negative (Proteobacteria) and gram-positive bacteria (Fir-



**Figure 2.** ARGs distribution and characterization: A total of 71 745 and 11 325 ARGs from blood and soil isolates, respectively, were identified after comparison with the Resfams database using the HMM model. The ARG-encoding proteins were classified according to their mechanism (A and B), their function described in the Resfams databases (C and D) and to the antibiotic that they confer resistance to (E and F). (A) Proportions and (B) value normalized by the number of genomes of the different mechanisms of resistance to antibiotics. (C) Proportions and (D) value normalized by the number of genomes of the different proteins related to antibiotic resistance. (E) Proportions and (F) value normalized by the number of genomes of the target antibiotics. Normalized values with significant differences in *t*-test for two samples (*P*-value < 0.05) were identified with the symbol \*.

micutes) (Supplementary Figure S2). Considering the efflux-based resistance category, 46.06 and 18.7% of the ARGs in Proteobacteria were from RND efflux and ABC transporter classes, respectively. On the other hand, in Firmicutes, these classes represented 3.42 and 40.91% of the ARGs, respectively. A similar proportion was found for soil genomes (Supplementary Figure S2). Searching for ARGs in the soil-isolated bacterial genomes, we verified that there are differences in the distribution of the transporters considering the blood and soil environments and the Proteobacteria and Firmicutes phyla. The efflux mechanisms identified as MFS efflux, SMR efflux and other efflux showed higher proportions in blood isolates. RND efflux showed similar proportions between soil and blood bacterial genomes while ABC transporter showed a higher proportion in soil bacterial genomes. However, when we consider the normalized values of ARGs per genome, blood-isolated bacterial genomes showed higher values in all transport and efflux mechanisms. In the specific case of Firmicutes and Proteobacteria, the comparative analysis with the soil-isolated

bacterial genomes showed that there is a small difference in proportion between Firmicutes and Proteobacteria of soil and blood in the ABC and RND transporters. Blood-isolated Firmicutes have a lower proportion of ABC transporters, but a higher proportion of MFS and RND transporters. Blood-isolated Proteobacteria, in turn, have a lower proportion of RND transporters, but a higher proportion of MFS, SMR and Other efflux (Supplementary Figure S2).

To specifically compare the distribution of ARGs in the genomes of blood and soil-isolated bacterial genomes, we evaluated the degree of identity between the proteins and Resfams, as well as the degree of identity between the proteins in the two samples to each other. The scores obtained for all proteins from blood and soil isolates were compared. We observed that the scores of all antibiotic resistance mechanisms of blood-isolated bacterial genomes are higher than those of soil isolates (Supplementary Figure S3). This result suggests that there is a higher identity among proteins from blood isolates in relation to Resfams than proteins from soil isolates. To specifically compare the

proteins from blood isolates to proteins from soil isolates in the ABC transporter and RND efflux we clustered all sequences. Through this analysis we note a clear distinction between blood and soil proteins in RND efflux and ABC transporter, with a little number of clusters containing proteins from both environments. Most of the groups were formed by proteins from the same environment. Overall, the clusters formed mainly by the blood isolates proteins were bigger than the clusters formed by the soil isolated proteins, however the soil groups presented a greater number of clusters (Supplementary Figure S4). Together, these two results suggest that, although we find similar proportions of these transporters in blood and soil-isolated bacterial genomes, these transporters correspond to different proteins.

We also found discrepancies between blood and soil-isolated bacterial genomes the two phyla in the target-based and antibiotic inactivation-based mechanisms category (Figure 2A and B). Despite the small number of genes identified within the target replacement ( $n = 564$ ) and target protection mechanisms ( $n = 232$ ), it was possible to determine that these genes were exclusive to the blood isolated Firmicutes and Proteobacteria phyla, respectively. The target-protection based mechanism was not identified in soil Firmicutes. ARGs within the antibiotic inactivation category were higher in proportion and in ARGs per genomes in the blood-isolated bacterial genomes (Figure 2A and B).

The distribution of resistance mechanisms within the first 10 genera with over 1000 identified ARGs each allowed us to verify the differences between gram-positive and gram-negative bacteria in more detail (Figure 3A and Supplementary Figure S5). Of these 10 genera, 6 were gram-negative bacteria and 4 were gram-positive bacteria. The genera of gram-negative bacteria had a very similar ARG profile, dominated by genes from the other antibiotic efflux, RND efflux and antibiotic inactivation categories. The genera of gram-positive bacteria, in turn, had two different profiles. *Staphylococcus* and *Enterococcus* presented a profile dominated by ABC transporters, target alterations and antibiotic inactivation genes. *Streptococcus* and *Listeria* presented a profile strongly dominated by ABC transporters with a smaller number of target alterations and antibiotic inactivation genes. Independence tests between antibiotic resistance mechanisms and gram type showed a significant dependence ( $P < 2.2e-16$ ) between these two factors (Figure 3A).

### Functional characterization of ARGs in blood isolates

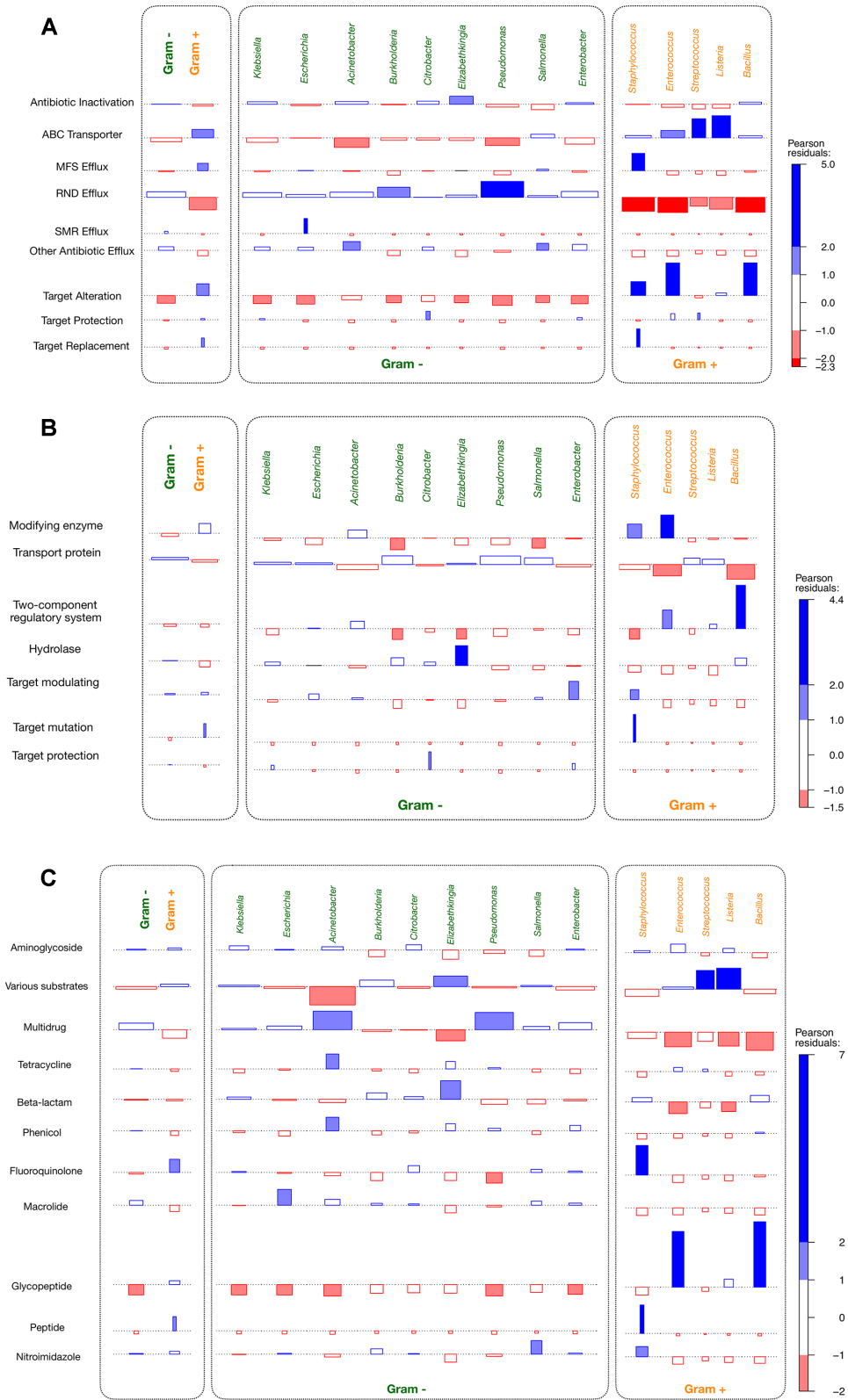
Using the functional categorization of Resfams results, we classified all ARGs by function (Figure 2C). The major protein-related function involved in antibiotic resistance was transport, with 64.74% of the sequences. Within this function were all types of transporters previously identified in this work (ABC, RDN, MFS, SMR and efflux pump). Following the transporters, we found the modifying enzymes, with 14% of the sequences, and hydrolases, with 9.47%. These values were close to those identified in the mechanisms of antibiotic inactivation. Interestingly, immediately after identifying hydrolases, we found two classes of proteins that, at first, did not have a canonical function in antibiotic resistance: gene modulating resistance pro-

teins (5.4%) and two-component regulatory system proteins (5.09%). The proportions of these functions, in general, are higher in soil genomes when compared to blood genomes. The normalized analysis by number of genomes showed that the amount of ARGs related to transport proteins, gene modulating and modifying enzymes is higher in blood genomes ( $P$ -value  $< 0.05$ ) (Figure 2D). As observed for the resistance mechanisms, we detected a significant functional discrepancy ( $P < 2.2e-16$ ) between the distribution of proteins from gram-positive and gram-negative bacteria (Figure 3B). The same dominance of transport proteins was observed at the genus level (Supplementary Figure S6).

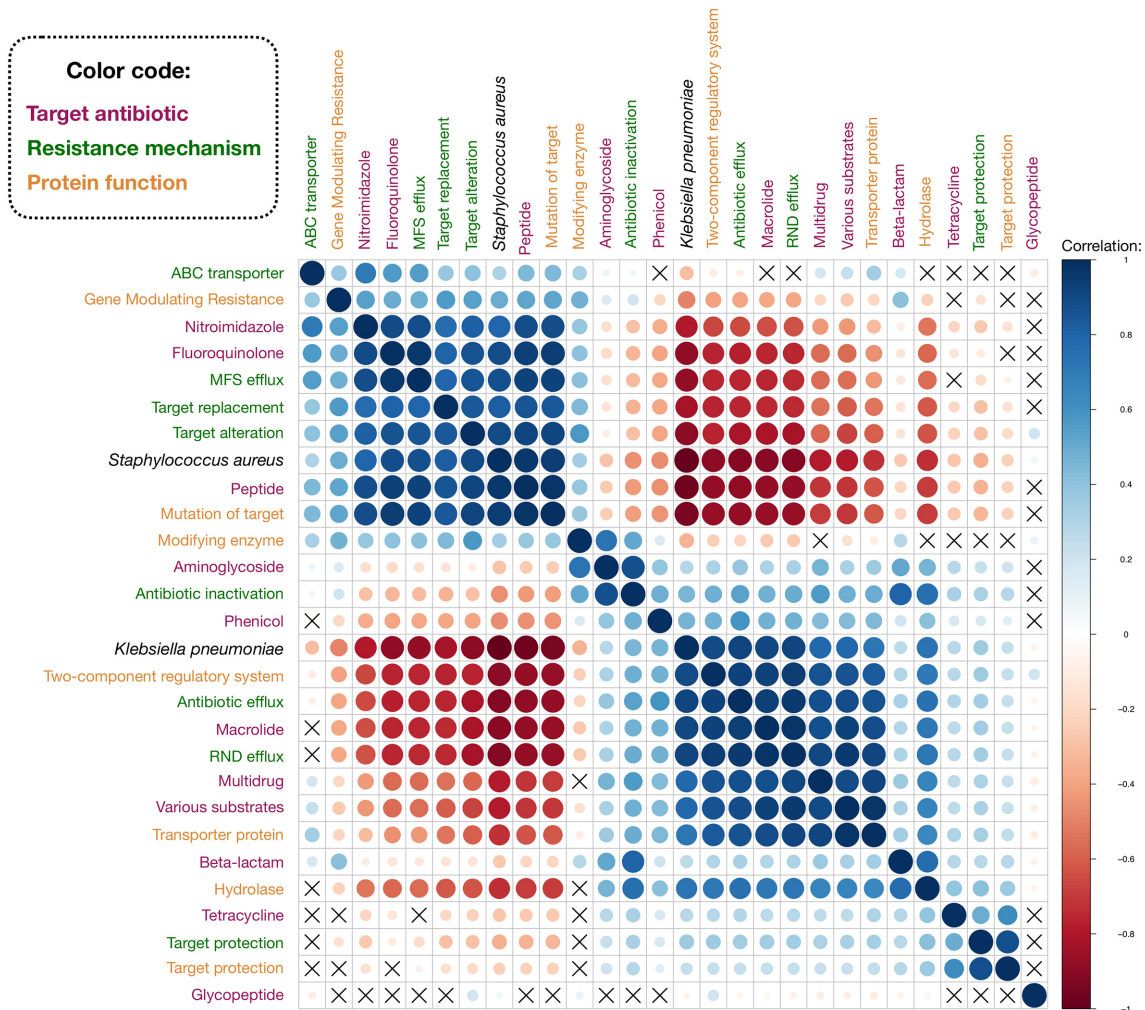
When we looked at the major classes of antibiotics to which these proteins confer resistance, proteins capable of attacking various classes of antibiotics (various substrates and multidrug) represented 58.38% of sequences (Figure 2E). Proteins that confer resistance to beta-lactam antibiotics appeared second (10.9%), followed by proteins that confer resistance to aminoglycosides (8.19%). The comparison of the proportions found for blood and soil-isolated bacterial genomes showed some discrepancies. The class various substrates, responsible for the transport of antibiotics and other molecules, is higher in soil genomes. The normalized analysis, however, showed similar values, but not significant. The multidrug class has a higher proportion in blood genomes when compared to soil genomes, both for proportions and for normalized analysis ( $P$ -value  $< 0.05$ ) (Figure 2F). Prediction of target antibiotics in the 10 genera with the highest number of identified genes showed, as in the previous analyses, a significant discrepancy ( $P < 2.2e-16$ ) between gram-positive and gram-negative bacterial genera (Figure 3C and Supplementary Figure S7).

### Correlation of ARG features of blood isolates.

We also tested the overall Pearson correlation between all types of antibiotic resistance mechanisms, protein functions and target antibiotics to identify possible coupled factors (Figure 4). Confirming our previous results, we found two very well-defined clusters: one related to *K. pneumoniae* (Kp-cluster) and the other related to *S. aureus* (Sa-cluster). Concerning the efflux mechanisms, we detected a positive correlation among *K. pneumoniae*, RND efflux, SMR efflux and other antibiotic efflux and a negative correlation with MFS efflux. For *S. aureus*, the opposite occurred: a strong positive correlation with ABC transporters and MFS efflux and a negative correlation for RND and other antibiotic efflux systems. ABC transporters are very general transporters. For this reason, the CARD database does not identify which antibiotics they provide resistance to and associates these proteins with 'various substrates' class more than 'multidrug'. In the target-based category, Kp had a weak correlation with target protection and antibiotic inactivation, while Sa had a positive correlation with target replacement and alteration. Within protein function, we found a strong correlation among Kp, transport protein, hydrolase and, interestingly, a two-component regulatory system. In the Sa cluster, we found a strong correlation with target mutations, modifying enzymes and gene modulation. Finally, for the antibiotic target, we found a strong positive correlation between Kp, various substrates, mul-



**Figure 3.** Association plot among antibiotic resistance features and taxonomy: association plot showing possible correlations between gram type or genera and the proportion of each resistance mechanism (A), protein function (B) and target antibiotics (C). Significant Pearson's residuals are shown as positive and negative whenever the genes in each category were enriched (blue) or depleted (red) compared to what would be expected by chance, respectively.



**Figure 4.** Correlation plot between *Klebsiella pneumoniae*, *Staphylococcus aureus* and all ARG features. The ARG occurrences in *K. pneumoniae* and *S. aureus* were distributed among antibiotic resistance mechanisms, protein functions and target antibiotics, and the correlation between terms was tested with the package *gplots* from R.

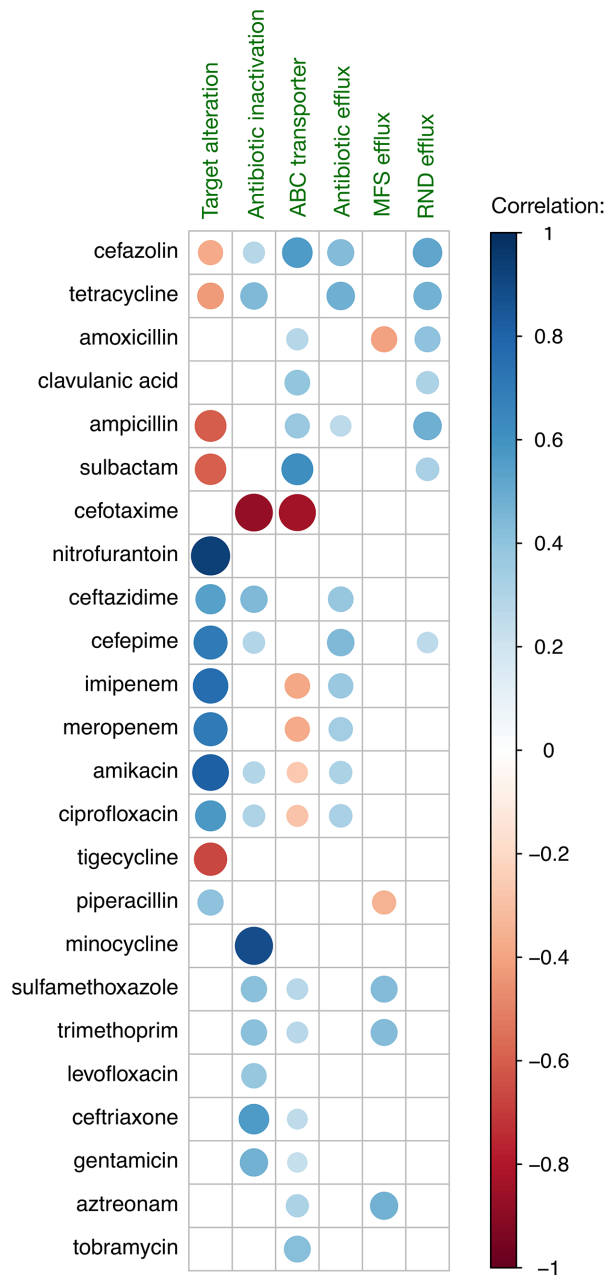
tidrug, Beta-lactam, aminoglycoside and macrolide. For *Sa*, we found a strong correlation with fluoroquinolones, nitroimidazoles and antimicrobial peptides. Based on these results, we can infer that the multiple resistance condition is a phenomenon more related to *K. pneumoniae* than *S. aureus*.

To check if there is a relationship between our results and experimentally measured phenotypic data, we identified, among the genomes of blood isolates, those who had information on the antibiotic to which they were resistant with the associated minimum inhibitory concentration (MIC) value. Of 3872 genomes, 84 had the necessary data. The correlation analysis between mechanisms and resistance phenotype revealed that the transporters are associated with almost all antibiotics for which experimental data were available (18 of 24,  $P < 0.05$ ). No relationship was found between transporters and cefotaxime, nitrofurantoin, tigecycline, minocycline, piperacillin and levofloxacin. Minocycline and levofloxacin correlate with antibiotic inactivation; nitrofurantoin and minocycline with target alteration; cefo-

taxime and tigecycline did not correlate with the analyzed mechanisms (Figure 5).

A strong correlation between beta-lactam and target alteration mechanism has also been identified. Although resistance to beta-lactam antibiotics has already been associated with the target alteration mechanism, the most common mechanism related to the beta-lactam antibiotic is enzymatic inactivation. To further investigate this correlation, we performed the same previous analysis for each species with five or more representatives with phenotypic resistance data. Only three species met the criteria: *E. coli* ( $n = 47$ ), *A. baumannii* ( $n = 16$ ) and *K. pneumoniae* ( $n = 5$ ). The correlation analysis carried out for the available blood-isolated *K. pneumoniae* genomes revealed that the main mechanism of resistance to beta-lactam is, indeed, antibiotic inactivation. Transport mechanisms were related to other class of antibiotics. Interestingly, these mechanisms were not complementary, that is, positive correlations for antibiotic inactivation had negative correlation for transporters for the same antibiotic (Supplementary Table S1). It was also pos-





**Figure 5.** Correlation plot between antibiotic resistance mechanisms and phenotypic features. Phenotypic antibiotic resistance (MIC) data were retrieved for 84 genomes of blood isolates and correlated with the ARGs identified in the respective genomes. Significant Pearson's residuals are shown as positive and negative whenever the genes in each category were enriched (blue) or depleted (red) compared to what would be expected by chance, respectively.

sible to verify in *K. pneumoniae* had a strong correlation between the antibiotics colistin and polymyxin B and the ABC transporter and RND efflux mechanisms. The same correlation was observed in *A. baumannii*. Cheng *et al.*, working with random insertion mutants of *K. pneumoniae* found that inactivating an RND efflux pump was able to increase sensitivity to colistin by 8-fold. We found that the correlation between carbapenems and target alteration was caused by the presence of genes related to this mechanism in all strains

of *A. baumannii* and in one strain of *E. coli*. For *E. coli*, the only strain capable of resisting concentrations of 16 mg/l of meropenem had one ARG related to target alteration. These features resulted in a strong positive correlation between meropenem and target alteration mechanism (Supplementary Table S1).

### Distribution of ARGs per genome per year

Among the two main phyla of bacteria isolated from blood (Proteobacteria and Firmicutes), it was possible to observe a large difference in the average number of ARGs per genome. While Proteobacteria (2177 genomes analyzed) averaged 23 ARGs per genome, Firmicutes (1594 genomes analyzed) averaged 12 ARGs per genome, a reduction of ~50%. For genus-level analysis, we limited the analysis to genera with at least 10 sequenced genomes (Table 1). Thus, the genus *Acinetobacter* appeared first with an average of 33 ARGs per genome. Within the phylum Proteobacteria, following the genus *Acinetobacter*, were the genera *Pseudomonas* (30 genes), *Escherichia* (26 genes), *Klebsiella* (26 genes), *Enterobacter* (25 genes) and *Serratia* (21 genes). Representatives of the Firmicutes phylum soon appeared with *Enterococcus* (21 genes) and *Staphylococcus* (18 genes). Interestingly, the next genus with the highest average of ARGs by genome was *Elizabethkingia*, from the phylum Bacteroides, with 17 genes. For comparison, the average number of ARGs per genomes in soil isolates was 12.9.

Information available in the PATRIC database also includes the year in which the samples were collected. From this information, it was possible to perform an ARG analysis over time (Figure 6). The earliest culture collection samples date from the early twentieth century, from the pre-antibiotic era and provide a rich source of information for understanding the evolution of the resistance phenomenon. Unfortunately, there are few records of this period for blood isolates. Overall, the number of ARGs per genome in blood isolates was very variable. However, a trend for the number of ARGs per genome to increase over time was observed (Figure 4). The number of deposited and analyzed genomes also increased over time, probably reflecting the evolution and decreased cost of DNA sequencing technologies, which contributes to more reliable analyses. The oldest *S. aureus* genome found with our search parameters dates back to 1963. When we plotted *S. aureus* data over time, we detected a large dispersion in the number of ARGs by genomes, and the values were close to the global average over the same period. In the case of *K. pneumoniae*, the oldest genome found dates back to 1997. This species shows a smaller dispersion, and the average of ARGs per genome was well above the global average over the same period (Figure 6).

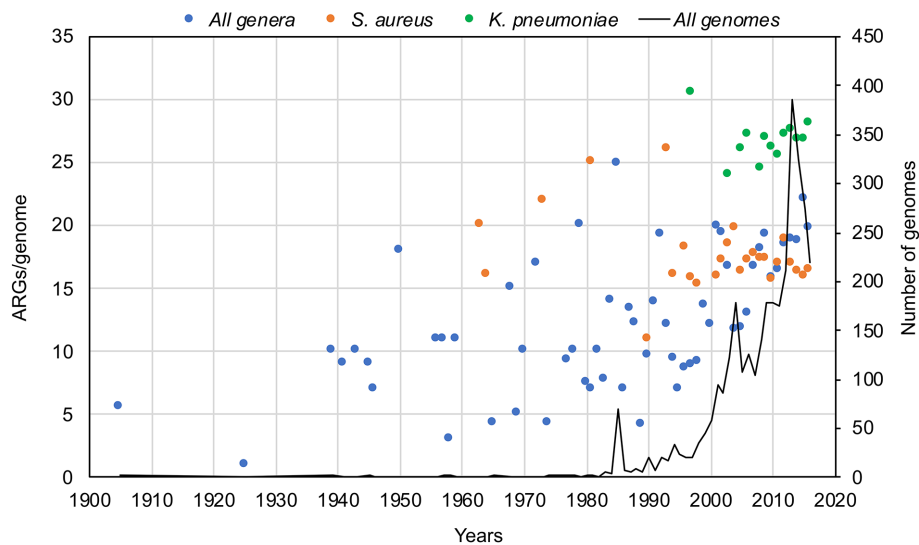
### DISCUSSION

Bacterial infections in the blood are a serious health problem due to their potential for rapidly developing sepsis. An additional complication to this condition arises when the bacteria causing this infection are resistant to antibiotics (24). The combination of these two factors constitutes one of the main threats to human health, as evidenced by the high mortality rate (24). In this paper, we conducted

**Table 1.** ARGs per genome

Phylum	Genus	#Genomes	#Total genes	#ARGs	ARGs/genomes	ARGs/total genes
Proteobacteria		2177	11 205 839	50 786	23.33	0.45
	<i>Acinetobacter</i>	226	881 493	7483	33.11	0.85
	<i>Pseudomonas</i>	212	1 429 075	6478	30.56	0.45
	<i>Escherichia</i>	422	2 230 964	11 342	26.88	0.51
	<i>Klebsiella</i>	588	3 493 634	15 667	26.64	0.45
	<i>Enterobacter</i>	60	322 658	1548	25.80	0.48
	<i>Serratia</i>	27	148 719	573	21.22	0.39
	<i>Burkholderia</i>	48	361 501	875	18.23	0.24
	<i>Stenotrophomonas</i>	24	117 04	416	17.33	0.36
	<i>Salmonella</i>	192	994 796	3124	16.27	0.31
	<i>Bordetella</i>	10	39 345	140	14.00	0.36
	<i>Yersinia</i>	28	124 305	361	12.89	0.29
	<i>Brucella</i>	39	136 372	339	8.69	0.25
	<i>Vibrio</i>	24	111 493	205	8.54	0.18
	<i>Bartonella</i>	25	37 315	149	5.96	0.40
	<i>Neisseria</i>	72	19 459	360	5.00	0.19
<i>Haemophilus</i>	26	49 726	130	5.00	0.26	
<i>Helicobacter</i>	34	82 958	65	1.91	0.08	
Firmicutes		1594	4 262 984	20 280	12.72	0.48
	<i>Enterococcus</i>	132	422 063	2770	20.98	0.66
	<i>Staphylococcus</i>	703	1 923 577	12 580	17.89	0.65
	<i>Listeria</i>	216	642 244	1950	9.03	0.30
	<i>Lactobacillus</i>	25	75 222	127	5.08	0.17
Bacteroides	<i>Streptococcus</i>	490	1 081 587	2424	4.95	0.22
		27	100 892	297	11.00	0.29
Actinobacteria	<i>Elizabethkingia</i>	10	39 325	175	17.50	0.45
	<i>Corynebacterium</i>	50	174 838	219	4.38	0.13
		18	45 239	62	3.44	0.14

All genera with more than 10 sequenced genomes were analyzed for the average number of ARGs per genome and the relation between ARGs and the total number of genes.



**Figure 6.** Historical distribution of the average number of ARGs in blood isolates. The average number of ARGs per year was obtained by dividing the total number of ARGs found in a given year by the total number of genomes available in the same year. The category ‘all genera’ does not include the genera *Staphylococcus* and *Klebsiella*. On the right axis, the total number of genomes, including *Staphylococcus* and *Klebsiella*, identified each year was plotted.

a broad analysis of publicly available data to profile taxonomic diversity, ARGs and resistance mechanisms in bacterial genomes isolated from human blood.

Taxonomic diversity analysis revealed that the two major bacterial phyla isolated from blood samples were Proteobacteria and Firmicutes, gram-negative and gram-positive bacteria, respectively. At the genus level, the main

genera listed were *Staphylococcus* and *Klebsiella* and, at the species level, *S. aureus* and *K. pneumoniae*. While infections caused by gram-positive bacteria were concentrated in the genus *Staphylococcus*, gram-negative bacteria, there was a broader distribution in different genera. *Staphylococcus* is a genus widely associated with blood infections, with *S. aureus* being the most related to clinical cases (25,26). *Staphy-*

*Staphylococcus aureus* is a commensal bacterium commonly found in the skin and upper airways and a pathogen responsible for ~20% of bloodstream infections (27,28). Other studies showed that *S. aureus* is the main bacterial species involved in sepsis (29). The emergence of methicillin-resistant *S. aureus* (MRSA), in turn, seriously aggravated sepsis, further increasing the mortality from blood infections caused by this bacterium (25). The classical mechanism of methicillin resistance in *S. aureus* occurs by the expression of an additional penicillin-binding protein (PBP2a) (30), but resistance by drug efflux also plays an important role, and several efflux pumps along with their transcriptional regulators have been identified (31). Similarly, the genus *Klebsiella* is widely associated with bloodstream infections. The main species involved in this context is *K. pneumoniae*, an enterobacteria naturally found in the environment on plant surfaces, soil and in the integument and mucosa of animals (32). In humans, this is a commensal organism that commonly inhabits the gastrointestinal microbiome (33). Historically, *K. pneumoniae* has been associated with nosocomial infections, being an opportunistic agent that primarily affects people undergoing treatment, immunocompromised patients, newborns and the elderly (34). A survey of the ICU found that *K. pneumoniae* was responsible for ~15% of bacterial infections between 1993 and 2004 (35). Since their identification, strains with high potential for virulence and/or resistance to antibiotics have been described, including carbapenemase-producing *K. pneumoniae* (KPC). The resistance of KPC is due to the production of enzymes that degrade broad-spectrum carbapenem antibiotics, a beta-lactam drug (36). The analysis of a database of soil-isolated bacterial genomes showed a more diverse taxonomic distribution with a very small proportion of *K. pneumoniae* and *S. aureus* indicating that there is, in fact, an enrichment for these species in cases of blood infection.

At the species level, we were able to detect bacteria of the group called 'ESKAPE' (formed by *Enterococcus faecium*, *S. aureus*, *K. pneumoniae*, *A. baumannii*, *P. aeruginosa* and *Enterobacter* species). This group of bacteria is strongly related to nosocomial infections and is becoming increasingly resistant and prevalent in hospital environments (37). In our analysis they stand out when considering the number of resistance genes per genome (Table 1), with at least 20 ARGs on average each (except for *S. aureus*). Other bacteria ranked higher in terms of the number of occurrences had less than 10 average ARGs per genome, as was the case for *Streptococcus pneumoniae*, *L. monocytogenes* and *Neisseria meningitidis*. This number is smaller than that found for soil-isolated bacterial genomes (12.9 ARGs per genomes). *Escherichia coli* is also highly prevalent in nosocomial infections (38) and resembles the other pathogens of this group, with an average of 26.9 resistance genes per genome. The *S. enterica* species also had a large number of occurrences in the sample ( $n = 192$ ), but it did not have as many ARGs as the ESKAPE nor as low as *Streptococcus*, *Listeria* and *Neisseria*, appearing to be intermediate among them.

In addition to the bacterial genera commonly found in blood infections, we also identified potential emerging bacteria that may become a problem in the future. Among these, one genus worth mentioning is *Elizabethkingia*. Unlike *Klebsiella* and *Staphylococcus*, the genus *Eliza-*

*bethkingia* belongs to the phylum Bacteroidetes, mostly consisting of environmental and commensal bacteria. This genus was created in 2005 with the transfer of the species *Chryseobacterium meningosepticum* and *Chryseobacterium miricola* (39). The former is known to cause nosocomial infections, especially in immunocompromised patients (40), whereas the latter was originally isolated from condensed water from the Russian space station (41) and rarely causes infections in humans (42). Later, *Elizabethkingia endophytica*, initially isolated from maize (43) and *Elizabethkingia anopheles*, isolated from *Anopheles gambiae* midgut (44), were added to this genus (45,46). Cases of bacterial infections of this genus are both hospital-acquired and community-acquired and usually have multidrug resistance phenotypes (45,47). In our study, we identified 10 genomes from the genus *Elizabethkingia* from blood isolates. In total representativeness, this number is not abnormal, but the number of ARGs per genome, 17 on average, was striking. This number is comparable to the average number of ARGs found in species such as *S. enterica* (16.39) and *S. aureus* (18.54), which are found in a large number of nosocomial infections. Resistance genes common to all genomes of this genus are related to antibiotic resistance to the phenicol and tetracycline classes of antibiotics and have at least four beta-lactam resistance genes and other multidrug-resistance efflux pump genes. The high incidence of beta-lactam resistance genes in *Elizabethkingia anophelis* has been described in other studies, which considered this species as a potential reservoir of new beta-lactamases (48).

Concerning the antibiotic resistance mechanisms, we found differences in the distribution of ARGs between gram-positive and gram-negative bacterial phyla. The target-based category had a prevalence of gram-positive bacteria. Most genes in this category belong to the class of target alteration genes in which the antibiotic target proteins are altered, preventing the antibiotic from binding to the protein. This modification can occur in two different ways, namely, amino acid point mutations in the protein (49) and enzymatic modification of the target antibiotic. Although a wide range of resistance is caused by point mutations in bacterial genes, the method used in this study for large scale analyzes, the HMM profiles, is not able to detect such mechanisms. Therefore, the influence, location and quantification of SNPs in ARGs has not been evaluated. On the other hand, it is possible to identify enzymes capable of target modification. An example of this case is the *emr* resistance gene (erythromycin ribosome methylation), which adds methyl groups to the adenine residue A2058 of the 23S rRNA, preventing macrolide antibiotics from binding to their site of action (50). We also found two resistance mechanisms specific to phyla, namely, target protection, found only in Proteobacteria, and target replacement, found only in Firmicutes. The target protection mechanism encodes proteins that remove the antibiotic from its target site or prevent its binding. Classic examples of this mechanism are the TetO proteins, which interact with domain IV of the 16S rRNA, changing the ribosome conformation and removing tetracycline from its target site, and TetM, which competes for the binding site of tetracycline (50,51). In the target replacement mechanism, the bacterium encodes proteins that have biochemical functions similar to those of proteins af-

ected by antibiotics, but are not affected by them (52). The antibiotic inactivation-based mechanism, in which antibiotics are degraded or modified, losing their action on cell, had a fairly balanced proportion of ARGs among the different phyla in blood isolates, but a large difference between blood and soil-isolated bacterial genomes.

Efflux systems were the most represented resistance mechanism in this dataset of blood isolates, as more than half of the ARGs of all phyla belonged to the different drug efflux class. The normalized comparative analysis between blood and soil genomes showed that there is an enrichment of ARGs related to antibiotic transport in the first group, mainly in RND efflux, MFS efflux, SMR efflux and other efflux proteins. Nevertheless, the efflux-based mechanisms RND and ABC transporters being found in large numbers in our analysis in both soil and blood-isolated bacteria, the proteins found in each group are completely different. Clustering of proteins classified as RND efflux and ABC transporters showed a clear division between blood and soil-isolated genomes. Clusters formed mostly by proteins from blood isolates are larger than those found for soil isolates proteins. This smaller diversity of blood-isolate proteins suggests a potential specialized function for this environment. Proteins from soil-isolated bacteria showed a larger number of clusters, however smaller in size, suggesting a higher diversity due to its complex environment. This feature also explains a reduced number of these genes in blood compared to the soil-isolated genomes. Although overshadowed by other resistance mechanisms, drug efflux has been shown to be an important element in the emergence of multi-resistant bacteria and has gained the interest of the scientific community in recent years. These mechanisms, which are also used for the elimination of other elements, such as metabolites and virulence factors, may also be specific for carrying an antibiotic, class of antibiotics or even various drugs. These efflux pumps are believed to act as a first line of bacterial defense, promoting antibiotic resistance at lower levels, but efficient for cell survival, allowing the expression of more stable resistance mechanisms capable of withstanding higher concentrations of antibiotics (31). Studies show that efflux-related genes undergo an increased gene expression due to the acquisition of plasmids with multiple resistance genes, even if the efflux coding gene were allocated on the chromosome (53). In addition, the overexpression of efflux pumps alone was related to multidrug resistance phenotypes (54) suggesting an importance of these genes in the multidrug resistance phenotype. Deletions and mutations in these genes also may result in an increase of susceptibility to various antibiotics like, beta-lactams, aminoglycosides, hospital-based disinfectants and even the last-resort antibiotic colistin (55,56). In *S. aureus*, several efflux proteins have already been described as being involved in antibiotic resistance mechanisms, such as those encoded by the *norABC* genes, which may confer resistance to fluoroquinolones and tetracyclines and *lmrS*, which is capable of eliminating various substances, including trimethoprim, erythromycin, kanamycin and phenicols (31). In *K. pneumoniae*, efflux genes such as *acrAB* are involved in increasing resistance to various antibiotics, such as fluoroquinolones, aminoglycosides and erythromycin (57). Furthermore, a recently discovered RND-efflux can increase the resistance to colistin up to 8-fold (56).

Despite the large dispersion of the average number of ARGs per genome as a function of time, the number of ARGs from *K. pneumoniae* isolates was constant, and above the general average. The isolates of *S. aureus* also had a constant average number of ARGs per genome over time, although these data were more dispersed. However, the average number of ARGs in *S. aureus* genomes, which was at one point above the general average, fell below the average by the mid 2010s. Unfortunately, this was not because of a reduction in the number of ARGs in *S. aureus* but an increase in the gain of ARGs in other species. This general increase in the average number of ARGs began mainly in the 1940s, when penicillin was introduced into clinical treatments and could be a result of selective pressure caused by antibiotic use (58).

Understanding the resistance phenomenon more deeply will help optimize treatments, reduce the use of antimicrobials and their consequences, such as selection of resistant microorganisms, lessen the side effects on patients and lower the cost of treatment. Consequently, this study strongly contributes to the search for more assertive empirical antimicrobials directly impacting patient improvement and survival (8).

## CONCLUSION

In this work, we performed a meta-analysis of 3872 genomes of bacteria isolated from blood infections and showed that 0.45% of the total translatable genes encode antibiotic-resistant proteins. *Staphylococcus aureus* and *K. pneumoniae* were by far the most represented bacteria in the blood samples; however, there were differences in antibiotic resistance mechanisms of these two bacteria. The comparison between genomes of blood-isolated and soil-isolated bacteria revealed a wide difference between the two communities, both in taxonomic and functional levels. Even the ABC and RND transporters, which appear in similar proportions in the two groups, showed a high divergence of sequences. On the other hand, other proteins more closely related to antibiotic resistance were enriched in genomes of bacteria isolated from blood. Still, the amount of ARGs identified in soil bacterial genomes was notable (0.31% of all proteins). Currently there is a lot of research and investment in the area of artificial intelligence in order to develop algorithms capable of assisting doctors who work in point-of-care medicine to decide which antibiotic to use (59,60). Some of these algorithms are based on the pathogen genome sequence, called sequence-based diagnosis. An important point on the way to achieving this goal is the availability of more genomes with reliable phenotypic resistance data. We believe that understanding the particularities of the host-pathogen relationship and the prevalent resistances associated with infections in specific sites of the body can help designing more efficient treatment strategies for each type of infection, saving time in treatment and increasing the patient survival rates.

## DATA AVAILABILITY

PATRIC, the Pathosystems Resource Integration Center, provides integrated data and analysis tools to support

biomedical research on bacterial infectious diseases. <https://www.patricbrc.org/>

Resfams is a curated database of protein families and associated profile hidden Markov models (HMMs), confirmed for antibiotic resistance function and organized by ontology. <http://www.dantaslab.org/resfams>

RefSoil is a curated reference database of sequenced genomes of organisms from the soil. The RefSoil genomes are a subset of NCBI's database of sequenced genomes, RefSeq (release 74), and have been manually screened to include only organisms that have previously been associated with soils. [https://github.com/germs-lab/ref\\_soil](https://github.com/germs-lab/ref_soil)

HMMER is used for searching sequence databases for sequence homologs, and for making sequence alignments. It implements methods using probabilistic models called profile HMMs. <http://hmmer.org/>

The scripts and data sets used during the development of this work are available at: [https://github.com/WillKlassen/Blood\\_resistome](https://github.com/WillKlassen/Blood_resistome)

The data set used in this study includes all genomes available in PATRIC version v3.5.34.

## SUPPLEMENTARY DATA

[Supplementary Data](#) are available at NARGAB Online.

## ACKNOWLEDGEMENTS

We thank the Coordination of Superior Level Staff Improvement (CAPES) and Carlos Chagas Institute (ICC), Fiocruz-PR, for their support.

## FUNDING

InovaFiocruz/Fundação Oswaldo Cruz [Grant number VPPCB-07-FIO-18-2-38]; CNPq [Grant number 424410/2018-4]

*Conflict of interest statement.* None declared.

## REFERENCES

- Sengupta,S., Chattopadhyay,M.K. and Grossart,H.P. (2013) The multifaceted roles of antibiotics and antibiotic resistance in nature. *Front. Microbiol.*, **4**, 47.
- Lewis,K. (2013) Platforms for antibiotic discovery. *Nat. Rev. Drug Discov.*, **12**, 371–387.
- Alekshun,M.N. and Levy,S.B. (2007) Molecular mechanisms of antibacterial multidrug resistance. *Cell*, **128**, 1037–1050.
- Martins,A., Hunyadi,A. and Amaral,L. (2013) Mechanisms of resistance in bacteria: an evolutionary approach. *Open Microbiol. J.*, **7**, 53–58.
- Bush,K. and Jacoby,G.A. (2010) Updated functional classification of beta-lactamases. *Antimicrob Agents Chemother.*, **54**, 969–976.
- Li,X.-Z. and Nikaido,H. (2009) Efflux-mediated drug resistance in bacteria: an update. *Drugs*, **69**, 1555–1623.
- Li,X.-Z., Plésiat,P. and Nikaido,H. (2015) The challenge of efflux-mediated antibiotic resistance in Gram-negative bacteria. *Clin. Microbiol. Rev.*, **28**, 337–418.
- Bookstaver,P.B., Nimmich,E.B., Smith,T.J., Justo,J.A., Kohn,J., Hammer,K.L., Troficanto,C., Albrecht,H.A. and Al-Hasan,M.N. (2017) Cumulative effect of an antimicrobial stewardship and rapid diagnostic testing bundle on early streamlining of antimicrobial therapy in Gram-negative bloodstream infections. *Antimicrob Agents Chemother.*, **61**, e00189-17.
- Fàbrega,A., Madurga,S., Giralt,E. and Vila,J. (2009) Mechanism of action of and resistance to quinolones. *Microb. Biotechnol.*, **2**, 40–61.
- Walsh,C. (2000) Molecular mechanisms that confer antibacterial drug resistance. *Nature*, **406**, 775–781.
- De Waele,J.J., Akova,M., Antonelli,M., Canton,R., Carlet,J., De Backer,D., Dimopoulos,G., Garnacho-Montero,J., Kesecioglu,J., Lipman,J. *et al.* (2018) Antimicrobial resistance and antibiotic stewardship programs in the ICU: insistence and persistence in the fight against resistance. A position statement from ESICM/ESCMID/WAAAR round table on multi-drug resistance. *Intensive Care Med.*, **44**, 189–196.
- Vincent,J.L., Rello,J., Marshall,J., Silva,E., Anzueto,A., Martin,C.D., Moreno,R., Lipman,J., Gomersall,C., Sakr,Y. *et al.* (2009) International study of the prevalence and outcomes of infection in intensive care units. *JAMA*, **302**, 2323–2329.
- Hampton,T. (2013) Report reveals scope of US antibiotic resistance threat. *JAMA*, **310**, 1661–1663.
- World Health Organization (WHO) (2014) In: *Antimicrobial resistance: global report on surveillance 2014*. WHO Press, Geneva.
- Wattam,A.R., Abraham,D., Dalay,O., Disz,T.L., Driscoll,T., Gabbard,J.L., Gillespie,J.J., Gough,R., Hix,D., Kenyon,R. *et al.* (2014) PATRIC, the bacterial bioinformatics database and analysis resource. *Nucleic Acids Res.*, **42**, D581–D591.
- Choi,J., Yang,F., Stepanauskas,R., Cardenas,E., Garoutte,A., Williams,R., Flater,J., Tiedje,J.M., Hofmocker,K.S., Gelder,B. *et al.* (2017) Strategies to improve reference databases for soil microbiomes. *ISME J.*, **11**, 829–834.
- Federhen,S. (2012) The NCBI Taxonomy database. *Nucleic Acids Res.*, **40**, D136–D143.
- Potter,S.C., Luciani,A., Eddy,S.R., Park,Y., Lopez,R. and Finn,R.D. (2018) HMMER web server: 2018 update. *Nucleic Acids Res.*, **46**, W200–W204.
- Gibson,M.K., Forsberg,K.J. and Dantas,G. (2015) Improved annotation of antibiotic resistance determinants reveals microbial resistomes cluster by ecology. *ISME J.*, **9**, 207–216.
- McArthur,A.G., Waglechner,N., Nizam,F., Yan,A., Azad,M.A., Baylay,A.J., Bhullar,K., Canova,M.J., De Pascale,G., Ejim,L. *et al.* (2013) The comprehensive antibiotic resistance database. *Antimicrob Agents Chemother.*, **57**, 3348–3357.
- Li,W. and Godzik,A. (2006) Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics*, **22**, 1658–1659.
- Warnes,G.R., Bolker,B., Bonebakker,L., Gentleman,R., Huber,W., Liaw,A., Lumley,T., Maechler,M., Magnusson,A. and Moeller,S. (2019) gplots: Various R Programming Tools for Plotting Data. pp. 1–68. <https://cran.r-project.org/web/packages/gplots/gplots.pdf>.
- Meyer,D., Zeileis,A. and Hornik,K. (2006) The strucplot framework: visualizing multi-way contingency tables with vcd. *J. Stat. Softw.*, **17**, doi:10.18637/jss.v017.i03.
- Borer,A., Saidel-Odes,L., Riesenberk,K., Eskira,S., Peled,N., Nativ,R., Schlaeffer,F. and Sherf,M. (2009) Attributable mortality rate for Carbapenem-Resistant klebsiella pneumoniae Bacteremia. *Infect. Control. Hosp. Epidemiol.*, **30**, 972–976.
- Wisplinghoff,H., Bischoff,T., Tallent,S.M., Seifert,H., Wenzel,R.P. and Edmond,M.B. (2004) Nosocomial bloodstream infections in US Hospitals: analysis of 24,179 cases from a prospective nationwide surveillance study. *Clin. Infect. Dis.*, **39**, 309–317.
- Marra,A.R., Camargo,L.F.A., Pignatari,A.C.C., Sukiennik,T., Behar,P.R.P., Medeiros,E.A.S., Ribeiro,J., Girão,E., Correa,L., Guerra,C. *et al.* (2011) Nosocomial bloodstream infections in Brazilian hospitals: Analysis of 2,563 cases from a prospective nationwide surveillance study. *J. Clin. Microbiol.*, **49**, 1866–1871.
- Lowy,F.D. (1998) Staphylococcus aureus infections. *N. Engl. J. Med.*, **339**, 520–532.
- Wertheim,H.F.L., Melles,D.C., Vos,M.C., Van Leeuwen,W., Van Belkum,A., Verbrugh,H.A. and Nouwen,J.L. (2005) The role of nasal carriage in Staphylococcus aureus infections. *Lancet Infect. Dis.*, **5**, 751–762.
- Mayr,F.B., Yende,S. and Angus,D.C. (2014) Epidemiology of severe sepsis. *Virulence*, **5**, 1–11.
- Beck,W.D., Berger-Bachi,B. and Kayser,F.H. (1986) Additional DNA in methicillin-resistant Staphylococcus aureus and molecular cloning of mec-specific DNA. *J. Bacteriol.*, **165**, 373–378.
- Costa,SS. (2013) Multidrug efflux pumps in Staphylococcus aureus: an update. *Open Microbiol. J.*, **7**, 59–71.

32. Podschun,R. and Ullmann,U. (1998) Klebsiella spp. as nosocomial pathogens: epidemiology, taxonomy, typing methods, and pathogenicity factors. *Clin. Microbiol. Rev.*, **11**, 589–603.
33. Wyres,K.L., Lam,M.M.C. and Holt,K.E. (2020) Population genomics of Klebsiella pneumoniae. *Nat Rev Microbiol.*, **18**, 344–359.
34. Bennett,J.E., John,E. and Dolin,R. (2014) In: *Mandell, Douglas, and Bennett's Principles and Practice of Infectious Diseases*. 8th Edition. Elsevier Saunders, Philadelphia.
35. Lockhart,S.R., Abramson,M.A., Beekmann,S.E., Gallagher,G., Riedel,S., Diekema,D.J., Quinn,J.P. and Doern,G.V. (2007) Antimicrobial resistance among Gram-negative bacilli causing infections in intensive care unit patients in the United States between 1993 and 2004. *J. Clin. Microbiol.*, **45**, 3352–3359.
36. Nordmann,P. and Poirel,L. (2002) Emerging carbapenemases in Gram-negative aerobes. *Clin. Microbiol. Infect.*, **8**, 321–331.
37. Rice,L.B. (2008) Federal funding for the study of antimicrobial resistance in nosocomial Pathogens: No ESKAPE. *J. Infect. Dis.*, **197**, 1079–1081.
38. Peleg,A.Y. and Hooper,DC. (2010) Hospital-acquired infections due to gram-negative bacteria. *N. Engl. J. Med.*, **362**, 1804–1813.
39. Kim,K.K., Kim,M.K., Lim,J.H., Park,H.Y. and Lee,S.T. (2005) Transfer of Chryseobacterium meningosepticum and Chryseobacterium miricola to Elizabethkingia gen. nov. as Elizabethkingia meningoseptica comb. nov. and Elizabethkingia miricola comb. nov. *Int. J. Syst. Evol. Microbiol.*, **55**, 1287–1293.
40. Bloch,K.C., Nadarajah,R. and Jacobs,R. (1997) Chryseobacterium meningosepticum: an emerging pathogen among immunocompromised adults. *Medicine (Baltimore)*, **76**, 30–41.
41. Li,Y., Kawamura,Y., Fujiwara,N., Naka,T., Liu,H., Huang,X., Kobayashib,K. and Ezakia,T. (2003) Chryseobacterium miricola sp. nov., a novel species isolated from condensation water of space station mir. *Syst. Appl. Microbiol.*, **26**, 523–528.
42. Green,O., Murray,P. and Gea-Banacloche,J.C. (2008) Sepsis caused by Elizabethkingia miricola successfully treated with tigecycline and levofloxacin. *Diagn. Microbiol. Infect. Dis.*, **62**, 430–432.
43. Kämpfer,P., Busse,H.J., McInroy,J.A. and Glaeser,S.P. (2015) Elizabethkingia endophytica sp. nov., isolated from Zea mays and emended description of Elizabethkingia anophelis Kämpfer et al. 2011. *Int. J. Syst. Evol. Microbiol.*, **65**, 2187–2193.
44. Kämpfer,P., Matthews,H., Glaeser,S.P., Martin,K., Lodders,N. and Faye,I. (2011) Elizabethkingia anophelis sp. nov., isolated from the midgut of the mosquito Anopheles gambiae. *Int. J. Syst. Evol. Microbiol.*, **61**, 2670–2675.
45. Lau,S.K.P., Chow,W.N., Foo,C.H., Curreem,S.O.T., Lo,G.C.S., Teng,J.L.L., Chen,J.H.K., Ng,R.H.Y., Wu,A.K.L., Cheung,I.Y.Y. et al. (2016) Elizabethkingia anophelis bacteremia is associated with clinically significant infections and high mortality. *Sci. Rep.*, **6**, 26045.
46. Perrin,A., Larssonneur,E., Nicholson,A.C., Edwards,D.J., Gundlach,K.M., Whitney,A.M., Gulvik,C.A., Bell,ME., Rendueles,O., Cury,J. et al. (2017) Evolutionary dynamics and genomic features of the Elizabethkingia anophelis 2015 to 2016 Wisconsin outbreak strain. *Nat. Commun.*, **8**, 15483.
47. Lau,S.K.P., Wu,A.K.L., Teng,J.L.L., Tse,H., Curreem,S.O.T., Tsui,S.K.W., Yi,H., Chen,J.H.K., Lee,R.A., Yuen,K.-Y. et al. (2015) Evidence for Elizabethkingia Anophelis transmission from mother to infant, Hong Kong. *Emerg. Infect. Dis.*, **21**, 232–241.
48. Campbell,E.A., Korzheva,N., Mustaev,A., Murakami,K., Nair,S., Goldfarb,A. and Darst,S.A. (2001) Structural mechanism for rifampicin inhibition of bacterial RNA polymerase. *Cell*, **104**, 901–912.
49. Weisblum,B. (1995) Erythromycin resistance by ribosome modification. *Antimicrob. Agents Chemother.*, **39**, 577–585.
50. Dönhöfer,A., Franckenberg,S., Wickles,S., Berninghausen,O., Beckmann,R. and Wilson,D.N. (2012) Structural basis for TetM-mediated tetracycline resistance. *Proc. Natl. Acad. Sci. U.S.A.*, **109**, 16900–16905.
51. Li,W., Atkinson,G.C., Thakor,N.S., Allas,U., Lu,C.C., Yan Chan,K., Tenson,T., Schulten,K., Wilson,KS., Haurlyuk,V. et al. (2013) Mechanism of tetracycline resistance by ribosomal protection protein Tet(O). *Nat. Commun.*, **4**, 1477.
52. Lim,D. and Strynadka,NCJ. (2002) Structural basis for the  $\beta$ -lactam resistance of PBP2a from methicillin-resistant Staphylococcus aureus. *Nat. Struct. Biol.*, **9**, 870–876.
53. Buckner,M.M.C., Saw,H.T.H., Osagie,R.N., McNally,A., Ricci,V., Wand,M.E., Woodford,N., Ivens,A., Webber,MA., Piddock,LJ.V. et al. (2018) Clinically relevant plasmid-host interactions indicate that transcriptional and not genomic modifications ameliorate fitness costs of Klebsiella pneumoniae carbapenemase-carrying plasmids. *MBio.*, **9**, e02303-17.
54. Yoon,E.J., Courvalin,P. and Grillot-Courvalin,C. (2013) RND-type efflux pumps in multidrug-resistant clinical isolates of Acinetobacter baumannii: major role for AdeABC overexpression and aders mutations. *Antimicrob Agents Chemother.*, **57**, 2989–2995.
55. Magiorakos,A., Srinivasan,A., Carey,R.B., Carmeli,Y., Falagas,M.E., Giske,C.G., Harbarth,S., Hindler,J.F., Kahlmeter,G., Olsson-Liljequist,B. et al. (2011) Multidrug-resistant, extensively drug-resistant and pandrug-resistant bacteria: an international expert proposal for interim standard definitions for acquired resistance. *Clin. Microbiol. Infect.*, **18**, 268–281.
56. Cheng,Y.H., Lin,T.L., Lin,Y.T. and Wang,J.T. (2018) A putative RND-type efflux pump, H239\_3064, contributes to colistin resistance through CrrB in Klebsiella pneumoniae. *J. Antimicrob Chemother.*, **73**, 1509–1516.
57. Padilla,E., Llobet,E., Doménech-Sánchez,A., Martínez-Martínez,L., Bengoechea,J.A. and Albertí,S. (2010) Klebsiella pneumoniae AcrAB efflux pump contributes to antimicrobial resistance and virulence. *Antimicrob Agents Chemother.*, **54**, 177–183.
58. Chain,E., Florey,H.W., Adelaide,M.B., Gardner,A.D., Oxford,D.M., Heatley,N.G., Jennings,M.A., Orr-Ewing,J. and Sanders,A.G. (1940) Penicillin as a chemotherapeutic agent. *Lancet*, **236**, 226–228.
59. Yelin,I., Snitser,O., Novich,G., Katz,R., Tal,O., Parizade,M., Chodick,G., Koren,G., Shalev,V., Kishony,R. et al. (2019) Personal clinical history predicts antibiotic resistance of urinary tract infections. *Nat. Med.*, **25**, 1143–1152.
60. Hicks,A.L., Wheeler,N., Sánchez-Busó,L., Rakeman,J.L., Harris,S.R. and Grad,Y.H. (2019) Evaluation of parameters affecting performance and reliability of machine learning-based antibiotic susceptibility testing from whole genome sequencing data. *PLoS Comput. Biol.*, **15**, e1007349.