

Combining functional and structural genomics to track antibiotic resistance genes in mobile elements in clinical bacterial strains

by

Tiago Cabral Borelli^{1§}, Gabriel Lencioni Lovate^{1§}, Ana Flavia Tonelli Scaranello¹, Lucas Ferreira Ribeiro¹, Livia Zaramela², Felipe Marcelo Pereira-dos-Santos³, María-Eugenia Guazzaroni^{1#} and Rafael Silva-Rocha^{3#}

¹Department of Biology, Faculdade de Filosofia, Ciências e Letras de Ribeirão Preto, University of São Paulo, São Paulo, SP, Brazil

²Department of Pediatrics, University of California San Diego, San Diego, California, USA

³Department of Cell and Molecular Biology, Faculdade de Medicina de Ribeirão Preto, University of São Paulo, São Paulo, SP, Brazil

[§]Both authors contributed equally to this work

[#]Both authors contributed equally to this work

Running Title: Tracking ARGs in clinical isolates

Keywords: Whole genome analysis, functional genomics, resistome, mobilome, antibiotic resistance genes.

*Correspondence to: Rafael Silva-Rocha
Faculdade de Medicina de Ribeirão Preto, Universidade de São Paulo
Av. Bandeirantes, 3.900. CEP: 14049-900.
Ribeirão Preto, São Paulo, Brazil
Tel.: +55 16 3602 3107; Fax: +55 16 3633 6840
E-mail: silvarocha@usp.br

Abstract

The rise of multi-antibiotics resistant bacteria represents an emergent threat to human health. Here, we investigate antibiotic resistance mechanisms in bacteria of several species isolated from an intensive care unit in Brazil. We used whole-genome analysis to identify antibiotic resistance genes (ARGs) and plasmids in 35 strains of Gram-negative and Gram-positive bacteria, including the first genomic description of *Morganella morganii* and *Ralstonia mannitolilytica* clinical isolates from South America. We identify a high abundance of beta-lactamase genes in highly resistant organisms, including seven extended-spectrum β -lactamases shared between organisms from different species. Additionally, we identify several ARGs-carrying plasmids indicating the potential for fast transmission of resistance mechanism between bacterial strains, comprising a novel IncFII plasmid recently introduced in Brazil from Asia. Through comparative genomic analysis, we demonstrate that some pathogens identified here are very distantly related to other bacteria isolated worldwide, demonstrating the potential existence of endemic bacterial pathogens in Brazil. Also, we uncovered at least two couples of (near)-identical plasmids exhibiting multi-drug resistance, suggesting that plasmids were transmitted between bacteria of the same or different species in the hospital studied. Finally, since many highly resistant strains carry several different ARGs, we used functional genomics to investigate which of them were indeed functional. In this sense, for three bacterial strains (*Escherichia coli*, *Klebsiella pneumoniae*, and *M. morganii*), we identify six beta-lactamase genes out of 15 predicted *in silico* as the main responsible for the resistance mechanisms observed, corroborating the existence of redundant resistance mechanisms in these organisms.

Importance

Big data and large-scale sequencing projects have revolutionized the field, achieving a greater understanding of ARGs identification and spreading at global level. However, given that microbiota and associated ARGs may fluctuate across geographic zones, hospital-associated infections within clinical units still remain underexplored in Brazil – the largest country in South America; 210 million inhabitants – and neighboring countries. This work highlighted the identification of several ARGs shared between species co-occurring simultaneously into a Brazilian hospital, some of them associated with large plasmids, mostly endowed with transposable elements. Also, genomic features of clinically underrepresented pathogens such *M. morganii* and *B. cepacia* were revealed. Taken together, our results demonstrate how structural and functional genomics can help to identify emerging mechanisms of shared antibiotic resistance in bacteria from clinical environments. Systematic studies as the one presented here should help to prevent outbreaks of novel multidrug resistance bacteria in healthcare facilities.

Introduction

Microbial resistance to antibiotics is a growing global concern. Established protocols in clinics to fight nosocomial infections include isolation of microorganisms from patient samples to allow its identification and to determine its antibiotic susceptibility (1). However, this process is time-consuming (48 h or more) and prone to pathogen misidentification (2). Therefore, the advent of next-generation sequencing (NGS) tools has allowed the rise of novel approaches to identify microbial pathogens and to fight infection (3). Thus, in the last two decades, whole-genome sequencing (WGS) of microbial pathogens has moved from being used as a basic research tool to understand pathogen's biology and evolution (4, 5), to an almost routine diagnostic tool to investigate outbreaks in hospitals and nosocomial infection pathways (6–9). For diagnostics purposes, current WGS technologies could even be cost effective for slow-growing pathogens such as *Mycobacterium tuberculosis*, providing faster and accurate results even for antibiotic resistance determination (10). Additionally, culture independent methods based on clinical metagenomics can be used to identify several pathogens from nucleic acids

extracted from patient samples without the need for microbial isolation (11–13). Furthermore, recent progress on the use of artificial intelligence tools has allowed the construction of computational models that can predict with high accuracy antimicrobial susceptibility of microbial pathogens based on WGS data (14–16).

While WGS analysis as routine for microbial identification is not a worldwide reality, it has been extensively used to investigate microorganisms' population structure at different scales. For example, Arias and coworkers using WGS analysis from 96 methicillin-resistant *Staphylococcus aureus* (MRSA) from 9 countries in Latin America demonstrated a high degree of variation in the genome of different isolates from those countries (17). The same study indicated that among sampled hospitals, those in Brazil presented a higher incidence of MRSA strains (up to 62%). Similarly, a recent work by David and coworkers investigated the path for the nosocomial spread of *Klebsiella pneumoniae* in 244 hospitals in 32 European countries (18). Using a well-defined sampling strategy and WGS analysis of more than 1700 *K. pneumoniae* strains, authors were able to quantify the role of intra-hospital pathogen dissemination, as well as some potential paths for the introduction of novel strains from the USA to Europe. In addition to those examples, large-scale WGS analysis has been used to investigate the molecular adaptation to different hosts, as in the work by Arimizu *et al.* in which authors analyzed *Escherichia coli* strains from human versus bovine samples (19).

Another key process playing a significant role in the rise of new microbial threats is the propagation of mobile virulence and antimicrobial resistance factors within these populations mediated especially by plasmids and transposons (5, 20). Therefore, the rapid evolution of plasmids through structural rearrangements, virulence genes acquisition, plasmid fusions, and propagation to pathogens can account for the fast dissemination of supervirulent or super-resistant bacteria (21–23). Understanding the very dynamic and complex processes could hold the potential to design new drugs aiming at reducing plasmid propagation between pathogens (24). While the use of WGS is currently growing worldwide, most studies (especially in Brazil) have been restricted to some particular species (such as *K. pneumoniae* or *S. aureus*) without considering their interplay with other species in hospital settings. Here, we investigate at the genomic level 35 strains from 18 different species (and 11 genera) isolated at the same two weeks in a reference public hospital in Brazil. WGS analysis indicates that many strains are only distantly related to those available at public databases. We aimed to identify antibiotic resistance genes (ARGs) and plasmids harboring these elements, as well as evidence for common resistance mechanisms shared between strains from the same or different species. We were able to identify several ARG-harboring plasmids, two of them present in both Gram-negative and Gram-positive strains, and seven beta-lactamases located in multiple hosts with 100% identity at the nucleotide level, two of which were inferred to be active using functional genomic library screening. In addition, comparative sequence analysis identified a novel IncFII *K. pneumoniae* plasmid harboring two ARGs, potentially indicating a recent introduction from Asia to Brazil.

Results and discussion

WGS analysis of clinical strains isolated from the same two weeks

We selected 35 bacterial strains isolated from different patient samples and performed WGS as represented in **Fig. 1**. While well-studied pathogens such as *K. pneumoniae*, *E. coli*, *P. aeruginosa* e *S. aureus* were well represented in the sampling, we were able to analyze pathogens with very few representative genomic information in March 2020. For example, two *Ralstonia mannitolilytica* strains were sequenced, but only nine complete or draft genomes were available at NCBI. Other underrepresented strains were *Streptococcus gallolyticus* (27 genomes available) and *Morganella morganii* (63 genomes). These numbers contrast with those from *K. pneumoniae*, *E. coli*, or *S. aureus*, where 8-20 thousand genome sequences are available. This evidence indicates that many clinically relevant pathogens have been underrepresented in WGS analysis efforts worldwide. For instance,

phylogenetic analysis of *Burkholderia cepacia* 540A against all 166 available genomes in NCBI (**Fig. S2**) indicates this strain is also very divergent but belongs to a branch formed by strains isolated from patients with cystic fibrosis in the UK and endophytic bacteria isolated from Australia (25). Therefore, some of the new WGSs generated here demonstrate a significant diversity of some underrepresented microorganisms and should serve as reference sequences for future studies on clinical isolates in Brazil and South America.

Identification of resistance genes in clinical strains

We next analyzed the existence of ARGs in the sequenced genomes. Analysis of ARGs using ARG-ANNOT database indicated a higher prevalence of beta-lactamase coding genes followed by amino-glycosidases (**Fig. 2A**), while analysis with DeepARG tool (26) showed the multidrug category as the most abundant on the genomes analyzed (**Fig. S3**). We next analyzed ARGs' distribution in the two most abundant groups, *E. coli* and *Klebsiella*. In the first case, each of the six *E. coli* strains displays a unique set of ARGs from different categories, and five particular beta-lactamases (*bla*_{TEM-105}, *bla*_{OXA-1}, *bla*_{KPC-2}, *bla*_{CTX-M-15}, and *bla*_{CMY-111}) were found in at least one genome (**Fig. 2B**). Only one strain (*E. coli* 456A) did not present any of these five *bla* genes, and this strain was sensitive to all antibiotics tested by Vitek 2 (**Fig. 2B/3**). For the *Klebsiella* group, we observed a much broader set of resistance markers (**Fig. 2C**), which was in accordance with the largest level of antibiotic resistance of this group compared to *E. coli*. Next, we investigated which beta-lactamase genes were conserved into the strains analyzed. For this, we compared all ~125 *bla* genes identified in **Fig. 2** and searched for those coding proteins with 100% identity at the amino acid sequence (**Fig. 3**). Using this approach, we identified seven *bla* genes (*bla*_{OXA-1}, *bla*_{OXA-10}, *bla*_{CTX-M-1}, *bla*_{KPC}, *bla*_{TEM}, *bla*_{HYDRO}, and *bla*_{BLP}) which have 100% aa identity between two or more strains. Notably, *bla*_{OXA-1}, *bla*_{CTX-M-1}, *bla*_{KPC}, and *bla*_{TEM} were present in 3 to 6 strains from different species. Interestingly, both *bla*_{OXA-1} and *bla*_{CTX-M-1} were found in *E. coli*, *K. pneumoniae* and *M. morgani*, which could reveal recent horizontal gene transfer between these strains. From these seven genes, 2 (*bla*_{OXA-1} and *bla*_{KPC}) were functional in the library screening presented below. Taken together, since the analyzed strains were isolated from hospitalized patients in the same two weeks, this evidence would indicate a recent mobilization of antibiotic resistance determinants either in the environment or in the hospital settings, as was reported previously (27–29).

Identification of ARGs located in plasmids

We next aimed to identify ARGs with potential mobilization through plasmids in the analyzed species. For this, we crossed the data from ARG-ANNOT with the prediction of plasmid elements generated by PlasmidFinder. Using this approach, we were able to identify nine potential plasmids from 7 species associated with at least one ARG. As shown in **Fig. 4A**, a ~42kb plasmid –(pKP98M3N42) – harboring a *bla*_{KPC-2} and a *sat-2A* resistance determinant were identified in *K. pneumoniae* 98M3, and this plasmid also carries transposases, recombinases, and type IV secretion system genes. An identical plasmid – (pKP125M3N44; 100% nucleotide sequence identity) – was also found in *K. pneumoniae* 125M3 (**Fig. 4A, Fig. S4A**). As mentioned before, this *bla*_{KPC-2} gene was identical in these two *K. pneumoniae* strains and in *E. coli* 126M3, but it was not located in a plasmid in the latter case (**Fig. 3**). In two recent studies, authors demonstrate ARG-carrying plasmids transfer in clinical settings between bacteria of the same or different species, which emphasize the significance of revealing potential plasmid-mediated outbreaks to efficiently track ARGs horizontal transmission in hospitals (30, 31).

Another strain, *K. pneumoniae* 508B, harbors two plasmids with ~136kb (pKP508BN15) and ~62kb (pKP508BN34). Both plasmids harbor transposon elements, with the larger one harboring a *sul2* resistance gene and the smaller one with two ARGs (*qnr-S1* and *bla*_{LAP-2}, **Fig. 4B-C**). In general, ARGs' coexistence with transposon elements was also observed for two plasmids identified in *E. coli* strains (**Fig. S4B-C**) and *Staphylococcus capitis* 732B (**Fig. S5A**). Finally, two almost identical small plasmids (~2.3kb; 99.94% nucleotide identity) were identified in *S. capitis* 732B and *S. epidermidis* 452B, which harbor an *aadC* and *ermC* resistance determinants (**Fig. S5B**). Taken together, these data demonstrate

the potential for dissemination of many ARGs genes identified here, and the existence of identical or near-identical plasmids between different species could indicate that these elements have been mobilizing among some of these species.

Experimental validation of functional beta-lactamases from highly-resistance bacteria

Once we distinguish several ARGs in the genomes analyzed, we decided to perform a functional screening to identify which of these genes could confer resistance into a heterologous host. For this, we selected three strains (*E. coli* 126M3, *K. pneumoniae* 508B, and *M. morgani* 538A) to construct genomic libraries into laboratory *E. coli* DH10B (Table 1). The libraries were constructed into the broad host range vector pSEVA232, which contains a kanamycin resistance marker, a broad host range *oriV* with a medium-copy number, and a variant of the *lacZα* multiple cloning site with a *Plac* promoter (32–34). In this way, the libraries generated during this study or individual plasmids of interest can be transferred to other bacterial strains for another screening or functional evaluation (**Fig. 5A**). The use of a medium copy-number plasmid allows a closer assessment of ARGs' natural genetic context in contrast to other studies that use high copy plasmids (35). Accordingly, plasmids with lower copy-number and monomeric states also tend to be more stably inherited throughout bacterial populations (36). We screened ~750.000 clones of each library against each antibiotic (amoxicillin, oxacillin, and penicillin G) and obtained 44 clones with unique sequences containing ARGs (**Fig. 5B** and **Table 2**).

Clones containing the *bla*_{KPC-2} were by far the more abundant in the screening (**Table 2** and **Fig. 6**). The genomic context of identified *bla*_{KPC-2} indicate that it is prone to suffer horizontal transfer, once it is flanked by transposases. Martínez *et al.* (37) describe a framework to prioritize the risk of ARGs, the Resistance Readiness Condition (RESCon). The RESCon algorithm considers the similarity of an ARG to known genes, functional evaluation, the clinical relevance of the antibiotic, presence of a mobile genetic element, and presence in a human pathogen. The characteristics of this ARG would categorize it as RESCon 1, an ARG with the highest possibility to thrive in a clinical setting (37). Although a framework better describing the impacts of gene transfer to prioritize risk is needed (38), the RESCon classification indicates clinical relevance of this ARG. With ARGs found in all 3 functional screenings, we were able to identify with high frequency six different *bla* genes using the functional approach presented here, being at least two of them present in plasmids – *bla*_{KPC-2} and *bla*_{LAP-2}; see next section – (**Fig. 6**).

We propose that functional genomics can be combined with current approaches based on large-scale sequencing in order to better understand the functional aspects of ARGs. Recent developments in machine learning (26) have provided tools to find ARGs that can be used to detect ARGs that would not be otherwise identified with sequence similarity tools. Still, the databases used to train those tools are biased for specific antibiotic resistance classes, such as beta-lactam, bacitracin, MLS, and efflux-pumps. Indeed, we could only find just a fraction of ARGs annotated through bioinformatics tools with our functional approach. This paradox could be due to the other ARGs not being functional in the experimental conditions used here or because our screening was not exhaustive enough to cover those sequences.

Mapping of plasmids in the Brazilian territory

To evaluate the biogeographic distribution of functional ARGs from *K. pneumoniae*, we performed genomic comparative analysis including 50 plasmids deposited in NCBI data bank and two of the plasmids identified in this study (pKP98M3N42, pKP508BN34). Distance-tree analysis showed the presence of two distinguishable groups. Plasmid pKP98M3N42 (identical to plasmid pKP125M3N44, also identified in this work) is located in the first group – comprising seven strains all reported in Brazilian cities from 2009 to 2015, but 1 in USA –, which should indicate that pKP98M3N42 is sharing structural features with the other plasmids positioned in this group (**Fig. 7A**). Additionally, pKP98M3N42 shares 100-99.9% sequence identity with IncX3 plasmids of these seven strains that have been previously related to play an important function in mediating horizontal transmission of *bla*_{KPC-2}

genes among hospital-associated members of the Enterobacteriaceae family (39, 40). Moreover, IncX3 plasmids carrying *bla*_{KPC-2} genes were also reported in countries very distant geographically, such as United States (41), Brazil (42, 43), Australia (44), Italy (45), France (46), South Korea (47, 48), China (49), Israel and Greece (39), to cite some. Interestingly, although IncX3 self-transmissible plasmids are widespread globally, the evident diversification of plasmids in the first separated branch of the tree (**Fig. 7A**) could indicate that plasmids from strains isolated in clinically relevant bacteria in Brazilian ground undergo their own structural reorganization.

On the other hand, distance tree representation of the 50 close plasmids' sequences related to pKP508BN34 did not show the presence of evident different groups. BLAST analysis showed that plasmid pKP508BN34 is widely distributed in *K. pneumoniae* strains, with 100-99,9% sequence identity and with query coverage ranging from 84% to 100% to the 18 closest related plasmids, some of them associated to hypervirulent strains – such as strain KP58, bearing plasmid pKP58-3 – (**Fig. 7B**). Mortality due to infection of *K. pneumoniae* was very rare in the past. However, this pathogen's fast evolution due to the gaining of hypervirulence plasmids allowed this bacterium to cause severe community-transmitted infections in relatively young and healthy hosts since the late 1980s (54, 55). pKP508BN34, with 63.067 bp in size and harboring various mobile elements that contain antimicrobial resistance genes, including *qnrS1* and *bla*_{LAP-2}, belongs to the IncFII plasmid group. As seen in the branch highlighted in **Fig. 7B**, most of the strains carrying the closest related plasmids to pKP508BN34 were described in different cities of China – but one in Germany, one in Japan, and three in Thailand –, and were associated with diverse host diseases, such as pneumoniae, pulmonary infection, urinary tract infection, intestinal infection, and diarrhea, according to the data deposited in the bioprojects of NCBI. In some cases, related strains were reported in asymptomatic patients (strains TH164 carrying plasmid pTH164-3 and strain TH114 bearing plasmid pTH114-3, both from Thailand), something reasonable since *K. pneumoniae* is also a member of the gut microbiota (50). To the best of our knowledge, this is the first time that this plasmid is reported in Brazil.

As shown in **Fig. 8A-B**, plasmids pKP98M3N42 and pKP508BN34 are highly structurally conserved between *K. pneumoniae* strains available in the databank. However, plasmid pKP508BN15 (which is present in *K. pneumoniae* 508B) presented a strong structural diversification in the region close to the antibiotic markers, and these changes seem to be related to the activity of the ISSpu21 transposon element located in this region (**Fig. 8C**). Interestingly, most related sequences available in the database are from strains worldwide including some isolated from Asia, Africa, North America, and Europe, but no example of sequences from South America.

Conclusions

Here, we sampled clinical bacterial strains to investigate the existence of ARGs and their association with mobile genetic elements. We could identify several ARGs shared between strains from different species, and some of these ARGs are associated with large plasmids, mostly endowed with transposable elements. Instead of focusing on strains from the same species or genus, our approach considered strains co-occurring simultaneously into a hospital setup, aiming to identify shared resistance mechanisms that could have been mobilizing in this environment. While our analysis does not provide unequivocal evidence that these resistance mechanisms are being mobilized among the strains analyzed, we found strongly conserved ARGs located in plasmids and associated with transposon elements that could represent potential mechanisms for the dissemination of antibiotic resistance among clinical strains.

Additionally, by using functional genomics, it was possible to investigate which of the ARGs candidates identified *in silico* could be the main factors associated with the resistance to beta-lactam antibiotics. Furthermore, for many of the bacterial species analyzed here, we could find a low number of available complete genome sequences at the NCBI database. Therefore, while thousand genome sequences are available for classical pathogens (*E. coli*, *K. pneumoniae*, *P. aeruginosa*, etc.), other clinically relevant pathogens such *M. morgani* and *B. cepacia* are underrepresented, which makes

challenging to track genomic events associated with the acquisition of pathogenicity elements or resistance mechanisms in hospital-associated infections. Finally, analysis of plasmids from *Klebsiella* strains allowed the identification of both well-known circulating variants in Brazil as new variants that seem to be recently acquired from Asia. Thus, we argue that more systematic efforts should be made to monitor the introduction and propagation of mobile genetic elements harboring ARGs, especially in South America, in order to avoid outbreaks of novel multidrug resistance bacteria.

Material and methods

Sample collection

Samples were taken at the Ribeirão Preto Clinics Hospital (HCRP, Ribeirão Preto, Brazil), a tertiary reference hospital in Latin America with 920 beds and 35,000 hospitalizations per year. Samples were isolated from different patients hospitalized at the weeks 44 and 48 of 2018. Strains were obtained from different samples, as indicated in **Fig. 1A**. Thirty-five strains were randomly selected from a total of 105 available and represent different species, which have a major prevalence for *Klebsiella* genus, *E. coli*, *Pseudomonas aeruginosa*, and the genus *Staphylococcus*. After strain characterization by Vitek 2, samples were inactivated and used for genomic DNA extraction and sequencing, as indicated below.

DNA extraction and genome sequencing

Total genomic DNA was extracted using Wizard Genomic DNA Purification Kit (Promega, Madison, WI, USA) following the manufacturer's instructions. **Fig. 1B** represents schematically the overall strategy used for WGS analysis. The DNA concentrations were measured fluorometrically (Qubit® 3.0, kit Qubit® dsDNA Broad Range Assay Kit, Life Technologies, Carlsbad, CA, USA). Purified DNA from 35 isolates was prepared for sequencing using Nextera XT DNA Library Prep Kit (Illumina). Libraries were quality assessed using 2100 Bioanalyzer (Agilent Genomics, Santa Clara, CA, USA) and subsequently sequenced using HiSeq 2x150 bp cycle kits (Illumina). On average, 5.5 million reads were generated per library. Adapters were trimmed using Trimmomatic v0.36. Samples were filtered of possible human contamination by aligning the trimmed reads against reference databases using Bowtie2 v2-2.2.3 with the following parameters (-D 20 -R 3 -N 1 -L 20 -very-sensitive-local). Overlapped reads were merged using Flash version 1.2.11. Merged and unmerged reads were assembled using Spades v3.12.0 with the following parameters (-k 21,33,55,77,99,127 --merge). Genome quality (completeness and contamination) was evaluated using CheckM v1.0.7 and QUAST. Genome annotations were performed using Prokka v1.11 with default parameters. Amino acid sequences of all genes identified using Prokka were aligned to the NDARO (National Database of Antibiotic Resistant Organisms) database obtained from NCBI (March, 2020). The alignment was performed using Diamond v0.8.24 with the following parameters (blastx -k 5 -f 6 -E value 0.001). Alignments with ≥ 60 similarity were selected for further analysis. Quality assessment of sequenced genomes is provided in **Fig. S1**. All genomes are available at NCBI under the BioProject number PRJNA641571.

Identification of antibiotic resistance genes, plasmids and phylogenomic analysis

The identification of antibiotic resistance genes was performed using the ABRicate pipeline by searching annotated genes using reference databases (ARG-ANNOT, NCBI AMRFinderPlus, CARD and ResFinder), as well as DeepARG (26). The identification of plasmid was performed using Plasmidfinder and contigs harboring ARGs and plasmid related genes were further analyzed in detail. For genomic comparison, reference and assembly genomic data were downloaded from NCBI databank. Phylogenomic analyses were performed using Parsnp and Gingr (51) and phylogenetic trees visualized using iTOL (52).

Genomic libraries construction

For the cloning of the genomic DNAs in the pSEVA232 (32–34) vector, 2 µg of genomic DNA from each strain were digested with Sau3AI. While, pSEVA232 digestion using BamHI and further dephosphorylation were performed. Genomic fragments from 1.5 to 6 kb and the linearized pSEVA232 vector were selected and incubated with the T4 DNA Ligase enzyme in a 2:1 insert/vector ratio. Then, ligations were transformed in the electrocompetent *E. coli* DH10B (53) with a MicroPulser electroporator (Bio-Rad - Hercules, USA). The resulting libraries were analyzed for the percentage of plasmids carrying genomic DNA and the average size of the insert they contained.

Determination of Minimum Inhibitory Concentrations (MICs)

The MICs were determined in the same culture medium (solid LB) and conditions of the screenings, employing serial dilution of the test antibiotics (Amoxicillin, Oxacillin, or Penicillin G). Solid medium plates supplemented with kanamycin (50 µg mL⁻¹), IPTG (100 µM) and dilutions of each beta-lactam antibiotic were inoculated with approximately 2.5 × 10⁶ CFUs of *E. coli* DH10B harboring the pSEVA232 vector. We performed dilutions of antibiotics and culture media according to the CLSI M100 supplement (54).

Screening and phenotype confirmation

Three pools of clones (2.5 × 10⁶ clones per plate) were plated from each library on solid LB supplemented with kanamycin (50 µg µL⁻¹), IPTG (100 µM) and inhibitory concentrations of each beta-lactam antibiotic (8 µg mL⁻¹ for amoxicillin, 32 µg mL⁻¹ for penicillin G or 256 µg mL⁻¹ for oxacillin). Positive clones were cultured in liquid LB medium supplemented with kanamycin (50 µg mL⁻¹) for extraction of plasmid DNA with the Wizard Plus SV Minipreps DNA Purification System Kit (Promega - Madison, USA). Plasmids with unique EcoRI/HindIII restriction patterns were subjected to re-transformation and phenotypic confirmation by streaking the clones in solid LB medium supplemented with kanamycin (50 µg mL⁻¹), IPTG (100 µM) and the test antibiotics.

Extraction of insert sequence from assembled genomes

Plasmid DNA was sequenced on both strands by primer walking using the ABI PRISMDye Terminator Cycle Sequencing Ready Reaction kit (PerkinElmer) and an ABI PRISM 377 sequencer (Perkin-Elmer) according to the manufacturer's instructions. A python script was developed to extract and annotate inserts from Sanger reads, which is available on GitHub. First, the algorithm converts .ab1 files to the fasta format and searches for the read sequence in the assembled genome using BLAST+ (55). We then extract the inter-read region to a fasta file, which is then annotated by PROKKA (56).

Acknowledgments

The authors are thanks to lab colleagues for their insightful comments and suggestions throughout the course of this study. We thank to Murilo Henrique Anzoline Cassiano for his initial support on genome assembly. We also thank to Dr. Roberto Martinez for providing the bacterial strains used in this study.

Funding

This work was supported by the São Paulo State Foundation (FAPESP, award # 2015/04309-1 and 2019/15675-0). TCB, GLL, LFR, MHAC and FMPS were supported by FAPESP fellowships (award # 2019/00390-0, 2018/18158-3, 2016/18827-7, 2019/06672-7 and 2018/1160-0).

References

1. Shetty N, Hill G, Ridgway GL. 1998. The Vitek analyser for routine bacterial identification and susceptibility testing: Protocols, problems, and pitfalls. *J Clin Pathol* <https://doi.org/10.1136/jcp.51.4.316>.

2. Pancholi P, Carroll KC, Buchan BW, Chan RC, Dhiman N, Ford B, Granato PA, Harrington AT, Hernandez DR, Humphries RM, Jindra MR, Ledebor NA, Miller SA, Brian Mochon A, Morgan MA, Patel R, Schreckenberger PC, Stamper PD, Simner PJ, Tucci NE, Zimmerman C, Wolk DM. 2018. Multicenter evaluation of the accelerate PhenoTest BC kit for rapid identification and phenotypic antimicrobial susceptibility testing using morphokinetic cellular analysis. *J Clin Microbiol* <https://doi.org/10.1128/JCM.01329-17>.
3. Köser CU, Ellington MJ, Cartwright EJP, Gillespie SH, Brown NM, Farrington M, Holden MTG, Dougan G, Bentley SD, Parkhill J, Peacock SJ. 2012. Routine Use of Microbial Whole Genome Sequencing in Diagnostic and Public Health Microbiology. *PLoS Pathog* <https://doi.org/10.1371/journal.ppat.1002824>.
4. Adams MD, Goglin K, Molyneaux N, Hujer KM, Lavender H, Jamison JJ, MacDonald IJ, Martin KM, Russo T, Campagnari AA, Hujer AM, Bonomo RA, Gill SR. 2008. Comparative genome sequence analysis of multidrug-resistant *Acinetobacter baumannii*. *J Bacteriol* <https://doi.org/10.1128/JB.00834-08>.
5. Wyres KL, Lam MMC, Holt KE. 2020. Population genomics of *Klebsiella pneumoniae*. *Nat Rev Microbiol*.
6. Snitkin ES, Won S, Pirani A, Lapp Z, Weinstein RA, Lolans K, Hayden MK. 2017. Integrated genomic and interfacility patient-transfer data reveal the transmission pathways of multidrug-resistant *Klebsiella pneumoniae* in a regional outbreak. *Sci Transl Med* 9.
7. Snitkin ES, Zelazny AM, Thomas PJ, Stock F, Henderson DK, Palmore TN, Segre JA. 2012. Tracking a hospital outbreak of carbapenem-resistant *Klebsiella pneumoniae* with whole-genome sequencing. *Sci Transl Med* <https://doi.org/10.1126/scitranslmed.3004129>.
8. Quick J, Cumley N, Wearn CM, Niebel M, Constantinidou C, Thomas CM, Pallen MJ, Moïemen NS, Bamford A, Oppenheim B, Loman NJ. 2014. Seeking the source of *Pseudomonas aeruginosa* infections in a recently opened hospital: An observational study using whole-genome sequencing. *BMJ Open* <https://doi.org/10.1136/bmjopen-2014-006278>.
9. Didelot X, Bowden R, Wilson DJ, Peto TEA, Crook DW. 2012. Transforming clinical microbiology with bacterial genome sequencing. *Nat Rev Genet*.
10. Pankhurst LJ, del Ojo Elias C, Votintseva AA, Walker TM, Cole K, Davies J, Fermont JM, Gascoyne-Binzi DM, Kohl TA, Kong C, Lemaitre N, Niemann S, Paul J, Rogers TR, Roycroft E, Smith EG, Supply P, Tang P, Wilcox MH, Wordsworth S, Wyllie D, Xu L, Crook DW. 2016. Rapid, comprehensive, and affordable mycobacterial diagnosis with whole-genome sequencing: A prospective study. *Lancet Respir Med* [https://doi.org/10.1016/S2213-2600\(15\)00466-X](https://doi.org/10.1016/S2213-2600(15)00466-X).
11. Chiu CY, Miller SA. 2019. Clinical metagenomics. *Nat Rev Genet*.
12. Naccache SN, Federman S, Veeraraghavan N, Zaharia M, Lee D, Samayoa E, Bouquet J, Greninger AL, Luk KC, Enge B, Wadford DA, Messenger SL, Genrich GL, Pellegrino K, Grard G, Leroy E, Schneider BS, Fair JN, Martínez MA, Isa P, Crump JA, DeRisi JL, Sittler T, Hackett J, Miller S, Chiu CY. 2014. A cloud-compatible bioinformatics pipeline for ultrarapid pathogen identification from next-generation sequencing of clinical samples. *Genome Res* <https://doi.org/10.1101/gr.171934.113>.
13. Greninger AL, Naccache SN, Federman S, Yu G, Mbala P, Bres V, Stryke D, Bouquet J, Somasekar S, Linnen JM, Dodd R, Mulembakani P, Schneider BS, Muyembe-Tamfum JJ, Stramer SL, Chiu CY. 2015. Rapid metagenomic identification of viral pathogens in clinical samples by real-time nanopore sequencing analysis. *Genome Med* <https://doi.org/10.1186/s13073-015-0220-9>.
14. Aytan-Aktug D, Clausen PTL, Bortolaia V, Aarestrup FM, Lund O. 2020. Prediction of Acquired Antimicrobial Resistance for Multiple Bacterial Species Using Neural Networks. *mSystems* <https://doi.org/10.1128/msystems.00774-19>.
15. Rahman SF, Olm MR, Morowitz MJ, Banfield JF. 2018. Machine Learning Leveraging Genomes from Metagenomes Identifies Influential Antibiotic Resistance Genes in the Infant Gut Microbiome. *mSystems* <https://doi.org/10.1128/msystems.00123-17>.
16. Macesic N, Bear Don't Walk OJ, Pe'er I, Tatonetti NP, Peleg AY, Uhlemann A-C. 2020. Predicting Phenotypic Polymyxin Resistance in *Klebsiella pneumoniae* through Machine Learning Analysis of Genomic Data. *mSystems* <https://doi.org/10.1128/msystems.00656-19>.
17. Arias CA, Reyes J, Carvajal LP, Rincon S, Diaz L, Panesso D, Ibarra G, Rios R, Munita JM,

- Salles MJ, Alvarez-Moreno C, Labarca J, Garcia C, Luna CM, Mejia-Villatoro C, Zurita J, Guzman-Blanco M, Rodriguez-Noriega E, Narechania A, Rojas LJ, Planet PJ, Weinstock GM, Gotuzzo E, Seas C. 2017. A prospective cohort multicenter study of molecular epidemiology and phylogenomics of *Staphylococcus aureus* bacteremia in nine Latin American countries. *Antimicrob Agents Chemother* <https://doi.org/10.1128/AAC.00816-17>.
18. David S, Reuter S, Harris SR, Glasner C, Feltwell T, Argimon S, Abudahab K, Goater R, Giani T, Errico G, Aspbury M, Sjunnebo S, Koraqi A, Lacey D, Apfalter P, Hartl R, Glupczynski Y, Huang TD, Strateva T, Marteva-Proevska Y, Andrasevic AT, Butic I, Pieridou-Bagatzouni D, Maikanti-Charalampous P, Hrabak J, Zemlickova H, Hammerum A, Jakobsen L, Ivanova M, Pavelkovich A, Jalava J, Österblad M, Dortet L, Vaux S, Kaase M, Gatermann SG, Vatopoulos A, Tryfinopoulou K, Tóth Á, Jánvári L, Boo TW, McGrath E, Carmeli Y, Adler A, Pantosti A, Monaco M, Raka L, Kurti A, Balode A, Saule M, Miciuleviciene J, Mierauskaite A, Perrin-Weniger M, Reichert P, Nestorova N, Debattista S, Mijovic G, Lopacic M, Samuelson Ø, Haldorsen B, Zabicka D, Literacka E, Caniça M, Manageiro V, Kaftandzieva A, Trajkovska-Dokic E, Damian M, Lixandru B, Jelesic Z, Trudic A, Niks M, Schreterova E, Pirs M, Cerar T, Oteo J, Aracil B, Giske C, Sjöström K, Gür D, Cakar A, Woodford N, Hopkins K, Wiuff C, Brown DJ, Feil EJ, Rossolini GM, Aanensen DM, Grundmann H. 2019. Epidemic of carbapenem-resistant *Klebsiella pneumoniae* in Europe is driven by nosocomial spread. *Nat Microbiol* <https://doi.org/10.1038/s41564-019-0492-8>.
 19. Arimizu Y, Kirino Y, Sato MP, Uno K, Sato T, Gotoh Y, Auvray F, Brugere H, Oswald E, Mainil JG, Anklam KS, Döpfer D, Yoshino S, Ooka T, Tanizawa Y, Nakamura Y, Iguchi A, Morita-Ishihara T, Ohnishi M, Akashi K, Hayashi T, Ogura Y. 2019. Large-scale genome analysis of bovine commensal *Escherichia coli* reveals that bovine-adapted *E. coli* lineages are serving as evolutionary sources of the emergence of human intestinal pathogenic strains. *Genome Res* <https://doi.org/10.1101/gr.249268.119>.
 20. Bennett PM. 2008. Plasmid encoded antibiotic resistance: Acquisition and transfer of antibiotic resistance genes in bacteria *British Journal of Pharmacology*.
 21. San Millan A. 2018. Evolution of Plasmid-Mediated Antibiotic Resistance in the Clinical Context. *Trends Microbiol*.
 22. Lermiaux NA, Cameron ADS. 2019. Horizontal transfer of antibiotic resistance genes in clinical environments. *Can J Microbiol* <https://doi.org/10.1139/cjm-2018-0275>.
 23. Porse A, Schönning K, Munck C, Sommer MOA. 2016. Survival and Evolution of a Large Multidrug Resistance Plasmid in New Clinical Bacterial Hosts. *Mol Biol Evol* <https://doi.org/10.1093/molbev/msw163>.
 24. Buckner MMC, Laura Ciusa M, Meek RW, Moorey AR, McCallum GE, Prentice EL, Reid JP, Alderwick LJ, Di Maio A, Piddock LJV. 2020. HIV drugs inhibit transfer of plasmids carrying extended-spectrum β -lactamase and carbapenemase genes. *MBio* <https://doi.org/10.1128/mBio.03355-19>.
 25. Johnson SL, Bishop-Lilly KA, Ladner JT, Daligault HE, Davenport KW, Jaissle J, Frey KG, Koroleva GI, Bruce DC, Coyne SR, Broomall SM, Li PE, Teshima H, Gibbons HS, Palacios GF, Rosenzweig CN, Redden CL, Xu Y, Minogue TD, Chain PS. 2016. Complete genome sequences for 59 *Burkholderia* isolates, both pathogenic and near neighbor. *Genome Announc* <https://doi.org/10.1128/genomeA.00159-15>.
 26. Arango-Argoty G, Garner E, Pruden A, Heath LS, Vikesland P, Zhang L. 2018. DeepARG: A deep learning approach for predicting antibiotic resistance genes from metagenomic data. *Microbiome* <https://doi.org/10.1186/s40168-018-0401-z>.
 27. Doi Y, Adams-Haduch JM, Peleg AY, D'Agata EMC. 2012. The role of horizontal gene transfer in the dissemination of extended-spectrum beta-lactamase-producing *Escherichia coli* and *Klebsiella pneumoniae* isolates in an endemic setting. *Diagn Microbiol Infect Dis* <https://doi.org/10.1016/j.diagmicrobio.2012.05.020>.
 28. Wyres KL, Wick RR, Judd LM, Froumine R, Tokolyi A, Gorrie CL, Lam MMC, Duchêne S, Jenney A, Holt KE. 2019. Distinct evolutionary dynamics of horizontal gene transfer in drug resistant and virulent clones of *Klebsiella pneumoniae*. *PLoS Genet* <https://doi.org/10.1371/journal.pgen.1008114>.
 29. Goren MG, Carmeli Y, Schwaber MJ, Chmelnitsky I, Schechner V, Navon-Venezia S. 2010.

- Transfer of Carbapenem-resistant plasmid from *Klebsiella pneumoniae* ST258 to *Escherichia coli* in patient. *Emerg Infect Dis* <https://doi.org/10.3201/eid1606.091671>.
30. Peter S, Bosio M, Gross C, Bezdán D, Gutierrez J, Oberhettinger P, Liese J, Vogel W, Dörfel D, Berger L, Marschal M, Willmann M, Gut I, Gut M, Autenrieth I, Ossowski S. 2019. Tracking of antibiotic resistance transfer and rapid plasmid evolution in a hospital setting by Nanopore sequencing. *bioRxiv* <https://doi.org/10.1101/639609>.
 31. Prussing C, Snavely EA, Singh N, Lapierre P, Lasek-Nesselquist E, Mitchell K, Haas W, Owsiak R, Nazarian E, Musser KA. 2020. Nanopore MinION Sequencing Reveals Possible Transfer of blaKPC-2 Plasmid Across Bacterial Species in Two Healthcare Facilities. *Front Microbiol* <https://doi.org/10.3389/fmicb.2020.02007>.
 32. Silva-Rocha R, Martínez-García E, Calles B, Chavarria M, Arce-Rodríguez A, de Las Heras A, Paez-Espino AD, Durante-Rodríguez G, Kim J, Nickel PI, Platero R, de Lorenzo V. 2013. The Standard European Vector Architecture (SEVA): a coherent platform for the analysis and deployment of complex prokaryotic phenotypes. *Nucleic Acids Res* 2012/11/28. 41:D666-75.
 33. Martínez-García E, Aparicio T, Goñi-Moreno A, Fraile S, de Lorenzo V. 2015. SEVA 2.0: an update of the Standard European Vector Architecture for de-/re-construction of bacterial functionalities. *Nucleic Acids Res* 43:D1183–D1189.
 34. Martínez-García E, Goñi-Moreno A, Bartley B, McLaughlin J, Sánchez-Sampedro L, Pascual Del Pozo H, Prieto Hernández C, Marletta AS, De Lucrezia D, Sánchez-Fernández G, Fraile S, de Lorenzo V. 2020. SEVA 3.0: an update of the Standard European Vector Architecture for enabling portability of genetic constructs among diverse bacterial hosts. *Nucleic Acids Res* 48:D1164–D1170.
 35. Rozwandowicz M, Brouwer MSM, Fischer J, Wagenaar JA, Gonzalez-Zorn B, Guerra B, Mevius DJ, Hordijk J. 2018. Plasmids carrying antimicrobial resistance genes in Enterobacteriaceae. *J Antimicrob Chemother* 73:1121–1137.
 36. Wein T, Hülter NF, Mizrahi I, Dagan T. 2019. Emergence of plasmid stability under non-selective conditions maintains antibiotic resistance. *Nat Commun* 10:1–13.
 37. Martínez JL, Coque TM, Baquero F. 2015. What is a resistance gene? Ranking risk in resistomes. *Nat Rev Microbiol* 13:116–123.
 38. Bengtsson-Palme J, Larsson DGJ. 2015. Antibiotic resistance genes in the environment: Prioritizing risks. *Nat Rev Microbiol* 13:396.
 39. Nordmann P, Naas T, Poirel L. 2011. Global spread of carbapenemase producing Enterobacteriaceae. *Emerg Infect Dis* <https://doi.org/10.3201/eid1710.110655>.
 40. Liakopoulos A, Van Der Goot J, Bossers A, Betts J, Brouwer MSM, Kant A, Smith H, Ceccarelli D, Mevius D. 2018. Genomic and functional characterisation of IncX3 plasmids encoding blaSHV-12 in *Escherichia coli* from human and animal origin. *Sci Rep* <https://doi.org/10.1038/s41598-018-26073-5>.
 41. Ho PL, Cheung YY, Lo WU, Li Z, Chow KH, Lin CH, Chan JFW, Cheng VCC. 2013. Molecular characterization of an atypical IncX3 plasmid pKPC-NY79 carrying blaKPC-2 in a *Klebsiella pneumoniae*. *Curr Microbiol* <https://doi.org/10.1007/s00284-013-0398-2>.
 42. Cerdeira LT, Cunha MPV, Francisco GR, Bueno MFC, Araujo BF, Ribas RM, Gontijo-Filho PP, Knöbl T, de Oliveira Garcia D, Lincopan N. 2017. IncX3 plasmid harboring a non-Tn4401 genetic element (NTEKPC) in a hospital-associated clone of KPC-2-producing *Klebsiella pneumoniae* ST340/CG258. *Diagn Microbiol Infect Dis* <https://doi.org/10.1016/j.diagmicrobio.2017.06.022>.
 43. Fuga B, Ferreira ML, Cerdeira LT, de Campos PA, Dias VL, Rossi I, Machado LG, Lincopan N, Gontijo-Filho PP, Ribas RM. 2020. Novel small IncX3 plasmid carrying the blaKPC-2 gene in high-risk *Klebsiella pneumoniae* ST11/CG258. *Diagn Microbiol Infect Dis* <https://doi.org/10.1016/j.diagmicrobio.2019.114900>.
 44. Partridge SR, Ginn AN, Wiklendt AM, Ellem J, Wong JSJ, Ingram P, Guy S, Garner S, Iredell JR. 2015. Emergence of blaKPC carbapenemase genes in Australia. *Int J Antimicrob Agents* <https://doi.org/10.1016/j.ijantimicag.2014.10.006>.
 45. Fortini D, Villa L, Feudi C, Pires J, Bonura C, Mammina C, Endimiani A, Carattoli A. 2016. Double copies of blaKPC-3::Tn4401a on an IncX3 plasmid in *Klebsiella pneumoniae* successful clone ST512 from Italy. *Antimicrob Agents Chemother* <https://doi.org/10.1128/AAC.01886-15>.

46. Kassis-Chikhani N, Frangeul L, Drieux L, Sengelin C, Jarlier V, Brisse S, Arlet G, Decré D. 2013. Complete nucleotide sequence of the first KPC-2- and SHV-12-encoding IncX plasmid, pKpS90, from *Klebsiella pneumoniae*. *Antimicrob Agents Chemother* <https://doi.org/10.1128/AAC.01712-12>.
47. Kim JO, Song SA, Yoon EJ, Shin JH, Lee H, Jeong SH, Lee K. 2017. Outbreak of KPC-2-producing Enterobacteriaceae caused by clonal dissemination of *Klebsiella pneumoniae* ST307 carrying an IncX3-type plasmid harboring a truncated Tn4401a. *Diagn Microbiol Infect Dis* <https://doi.org/10.1016/j.diagmicrobio.2016.12.012>.
48. Jeong S, Kim JO, Yoon EJ, Bae IK, Lee W, Lee H, Park Y, Lee K, Jeong SH. 2018. Extensively drug-resistant *Escherichia coli* sequence type 1642 carrying an IncX3 plasmid containing the blaKPC-2 Gene associated with transposon Tn4401a. *Ann Lab Med* <https://doi.org/10.3343/alm.2018.38.1.17>.
49. Yang Q, Fang L, Fu Y, Du X, Shen Y, Yu Y. 2015. Dissemination of NDM-1-producing Enterobacteriaceae mediated by the IncX3-type plasmid. *PLoS One* <https://doi.org/10.1371/journal.pone.0129454>.
50. Acton QA. 2012. *Klebsiella pneumoniae*: New Insights for the Healthcare Professional: 2011 Edition: ScholarlyBrief. ScholarlyEditions.
51. Treangen TJ, Ondov BD, Koren S, Phillippy AM. 2014. The harvest suite for rapid core-genome alignment and visualization of thousands of intraspecific microbial genomes. *Genome Biol* <https://doi.org/10.1186/s13059-014-0524-x>.
52. Letunic I, Bork P. 2019. Interactive Tree Of Life (iTOL) v4: recent updates and new developments. *Nucleic Acids Res* <https://doi.org/10.1093/nar/gkz239>.
53. Drury L. 1996. Transformation of bacteria by electroporation. *Methods Mol Biol* 58:249–256.
54. CLSI. 2018. Performance Standards for Antimicrobial Susceptibility Testing 28th ed. CLSI supplement M100. Clinical and Laboratory Standards Institute, Wayne, PA.
55. Camacho C, Coulouris G, Avagyan V, Ma N, Papadopoulos J, Bealer K, Madden TL. 2009. BLAST+: Architecture and applications. *BMC Bioinformatics* <https://doi.org/10.1186/1471-2105-10-421>.
56. Seemann T. 2014. Prokka: rapid prokaryotic genome annotation. *Bioinformatics* 30:2068–2069.
57. Zulkower V, Rosser S. 2020. DNA Features Viewer: a sequence annotation formatting and plotting library for Python. *Bioinformatics* <https://doi.org/10.1093/bioinformatics/btaa213>.
58. Alikhan NF, Petty NK, Ben Zakour NL, Beatson SA. 2011. BLAST Ring Image Generator (BRIG): Simple prokaryote genome comparisons. *BMC Genomics* <https://doi.org/10.1186/1471-2164-12-402>.

Tables

Table 1 – Features of the generated genomic libraries.

Genomic library	<i>E. coli</i> 126M3	<i>K. pneumoniae</i> 508B	<i>M. morgani</i> 538A
Total number of clones	13442	25693	11804
Percentage of clones with insert (%)	90	90	90
Number of clones with insert	12122	23124	10624
Insert size variation (kb)	1.0 – 10.5	0.2 – 2.2	0.2 – 6.5
Average insert size (kb)	5.2	1.4	3.0
Total genomic library size (mb)	62.8	32.4	32.1
Estimated genome coverage	11.9 ×	5.8 ×	7.8 ×

Table 2 – Comparison of ARGs identification using *in silico* and functional approaches.

Strain	<i>In silico</i> (argannot)	<i>No. of sequenced resistant clones</i>
<i>E. coli</i> 126M3	Penicillin_Binding_Protein_Ecoli	-
	AmpC1_Ecoli	-
	<i>bla</i> _{KPC-2}	28
	AmpC2_Ecoli	1
	<i>bla</i> _{OXA-1}	-
	ampH_Ecoli	-
<i>K. pneumoniae</i> 508B	<i>bla</i> _{SHV-11}	-
	Penicillin_Binding_Protein_Ecoli	-
	<i>bla</i> _{LAP-2}	2
	ampH	-
	<i>bla</i> _{CTX-M-15}	8
	<i>bla</i> _{OXA-1}	1
<i>M. morgani</i> 538A	<i>bla</i> _{CTX-M-15}	-
	<i>bla</i> _{OXA-1}	-
	<i>bla</i> _{MOR-2}	4

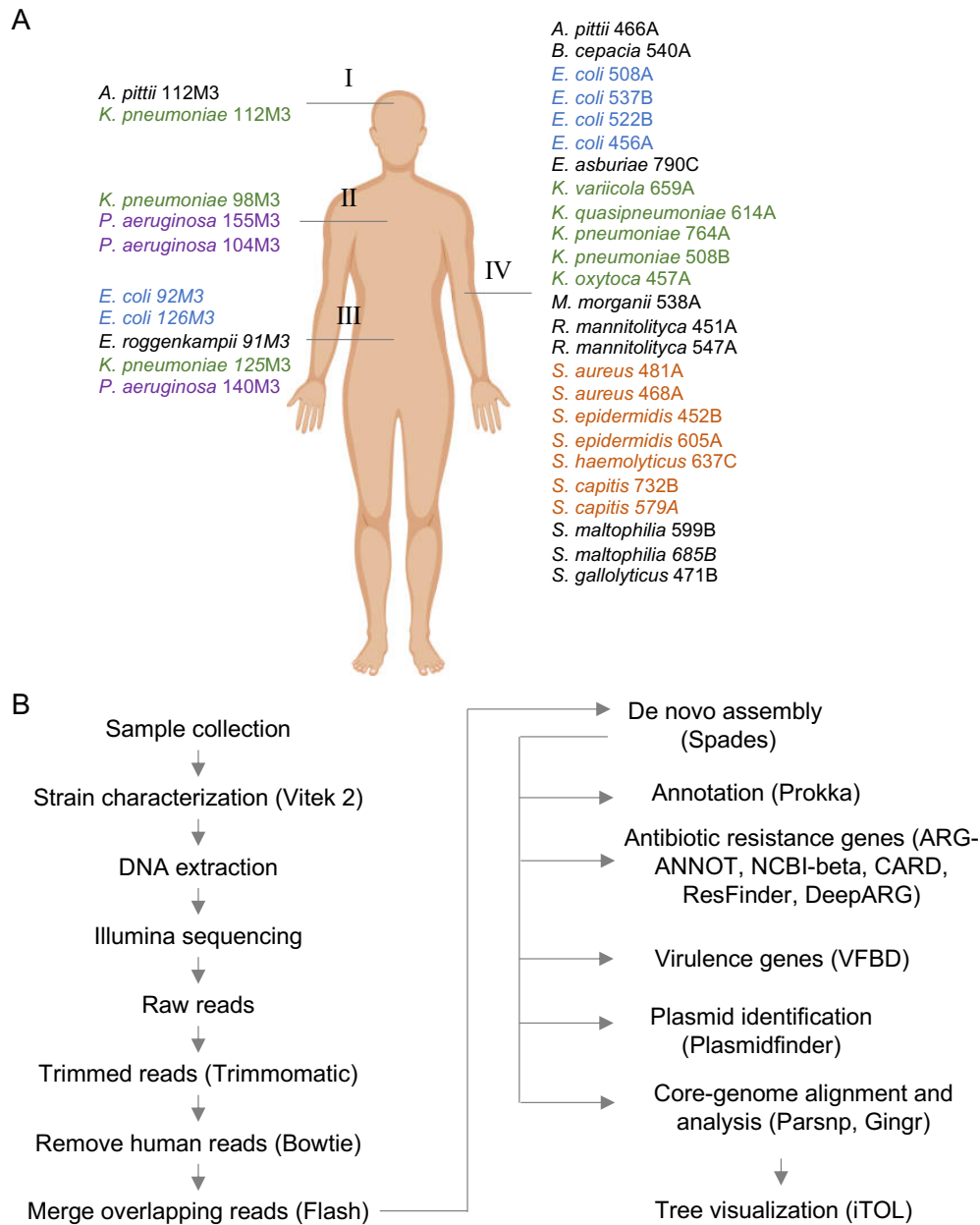


Figure 1. Overall strategy used for the whole-genome analysis of clinical strains. **A)** In total, 35 bacterial strains were isolated from several samples, such as cerebrospinal fluid (I, 2 strains), bronchoalveolar lavage (II, 3 strains), ascitic fluid (III, 5 strains), and blood (IV, 25 strains). The four most common bacterial groups (*K. pneumoniae*, *E. coli*, *P. aeruginosa*, and *S. genus*) are colored. **B)** Schematic representation of the main bioinformatic pipeline used for genome sequencing, assembly, annotation, identification of ARGs, and virulence factors and genomics analysis.

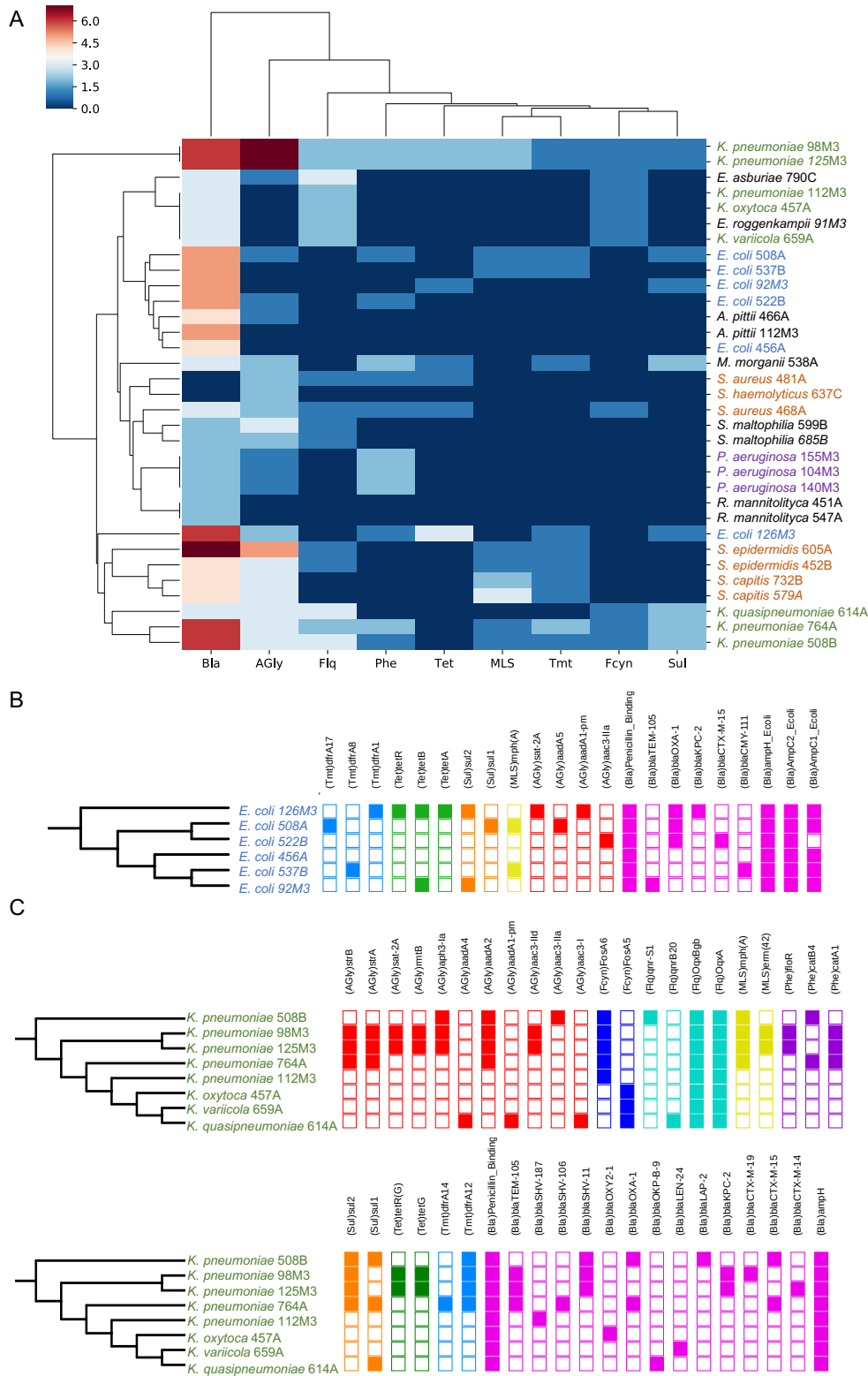


Figure 2. Identification of ARGs in nearly sequenced genomes. **A)** Heatmap showing the presence of ARGs identified by ARG-ANNOT. Only *B. cepacia* did not present any identified ARG. Data was clustered using hierarchical mapping with Euclidian distance. Blue to red scale indicate number of ARG for each strain in each category, as indicated in the legend. **B)** Distribution of different ARGs per genome of *E. coli*, colored by antibiotic category. The maximum likelihood phylogeny for the strains was based on core-genome. **C)** Distribution of different ARGs per genome of *K. pneumoniae*, following the scheme in B.

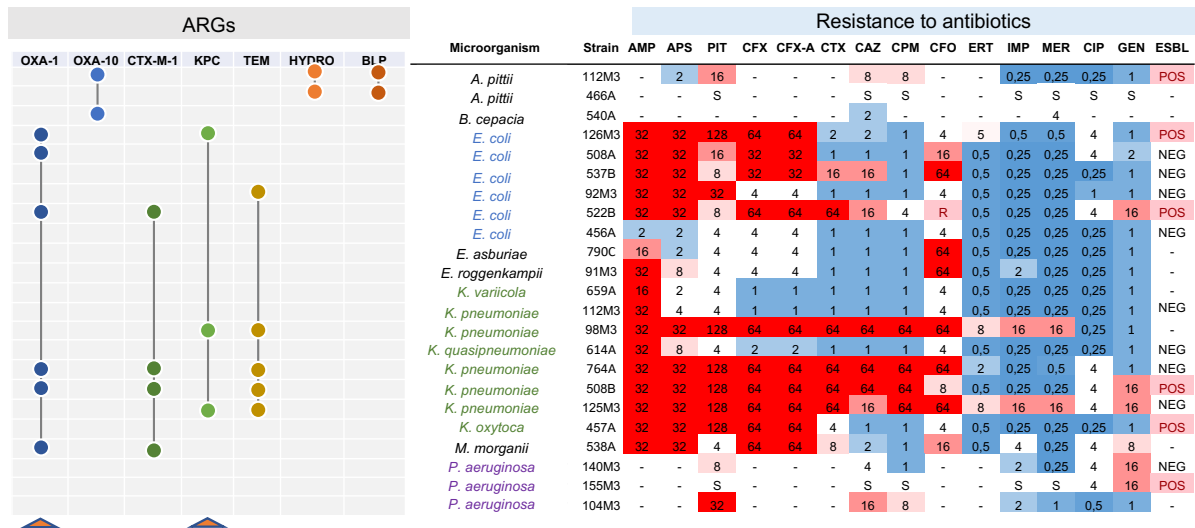


Figure 3. Shared beta-lactamases with 100% identity at the protein level. Seven beta-lactamase coding genes were found as shared among gram-negative strains analyzed. Connected circles indicate that the genes are presented in those strains. On the right, the antibiotic resistance profile of the analyzed strains. Genes for *bla*_{OXA-1} and *bla*_{KPC} are highlighted (red triangles) since these genes were identified in the functional screening carried out in this study. On the right, antibiotic resistance levels are indicated, with numbers indicating the resistance levels in µg/mL. Red indicates resistance to the antibiotic while blue denotes sensibility. The final column indicated if the strain is ESBL positive or negative.

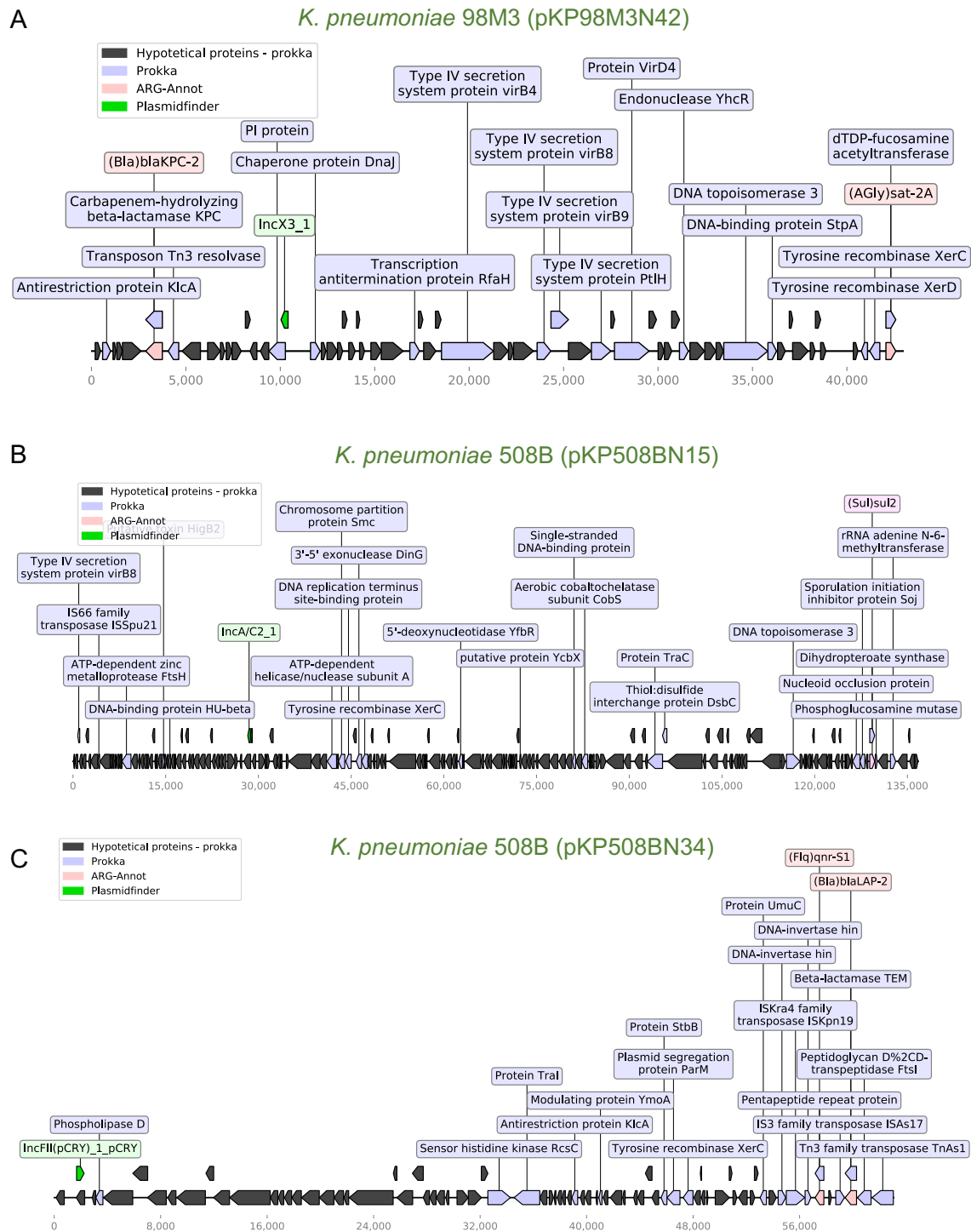


Figure 4. Schematic representation of genes of three *K. pneumoniae* plasmids. **A)** Plasmid pKP98M3N42 (43 kb) from *K. pneumoniae* 98M3. This plasmid carries two ARGs (*bla*_{KPC-2} and *sat-2A*), elements of a type IV secretion system, two resolvases, and a Tn3 transposase. This plasmid is very similar to pKP125M3N44 from *K. pneumoniae* 125M3 (Fig. S4A). Whole-plasmid visualization was performed using a python module for prokaryotic genome analyses – DnaFeaturesView – and matplotlib module combined, as well as ARG-ANNOT and Prokka's results (57). **B)** Plasmid pKP508BN15 (136.8 kb) from *K. pneumoniae* 508B transports a *sul2* and an rRNA adenine n-6-methyltransferase (*ermC*) resistance markers. A transposase, a *xerC* recombinase, and a type IV secretion system *virB8* protein are also found in this plasmid. **C)** Plasmid pKP508BN34 (63 kb) is also from *K. pneumoniae* 508B and carries several transposases and two resistance determinants, *qnr-S1* and *bla*_{LAP-2}. Legends represent the colors code for the identified genes.

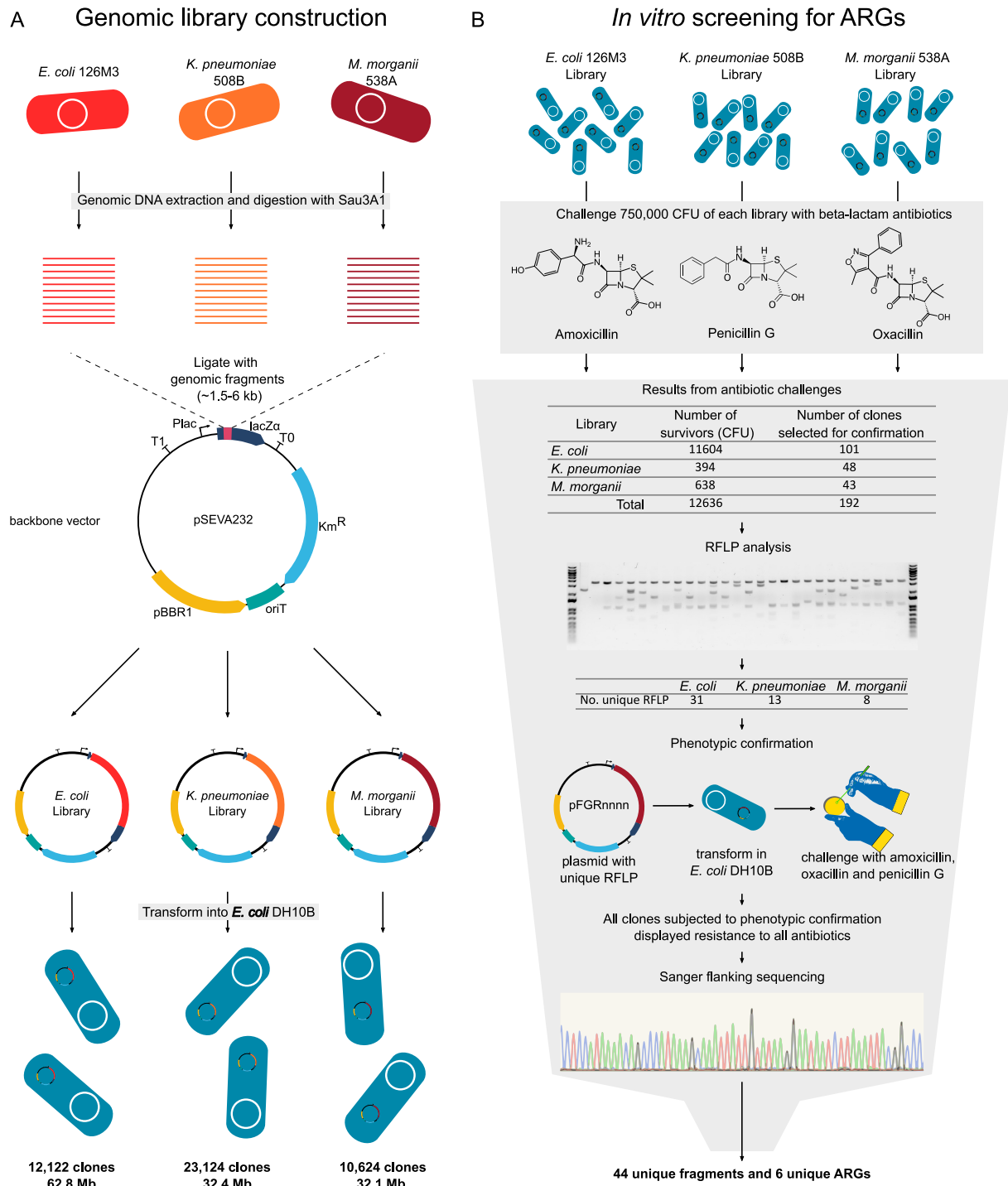


Figure 5. Experimental design for functional genomics analysis. **A)** Strains and backbone used to construct the library and features of the obtained libraries. **B)** Functional screening for beta-lactam resistant clones, phenotypic confirmation, and sequence identification. CFU – colony forming unit, RFLP - restriction fragment length polymorphism, KmR - kanamycin resistance marker, pBBR1 - a broad host range *oriV*, *lacZα* - *lacZα* gene with multiple cloning site, *Plac* - *Plac* promoter, pFGRnnnn – plasmid naming schema.

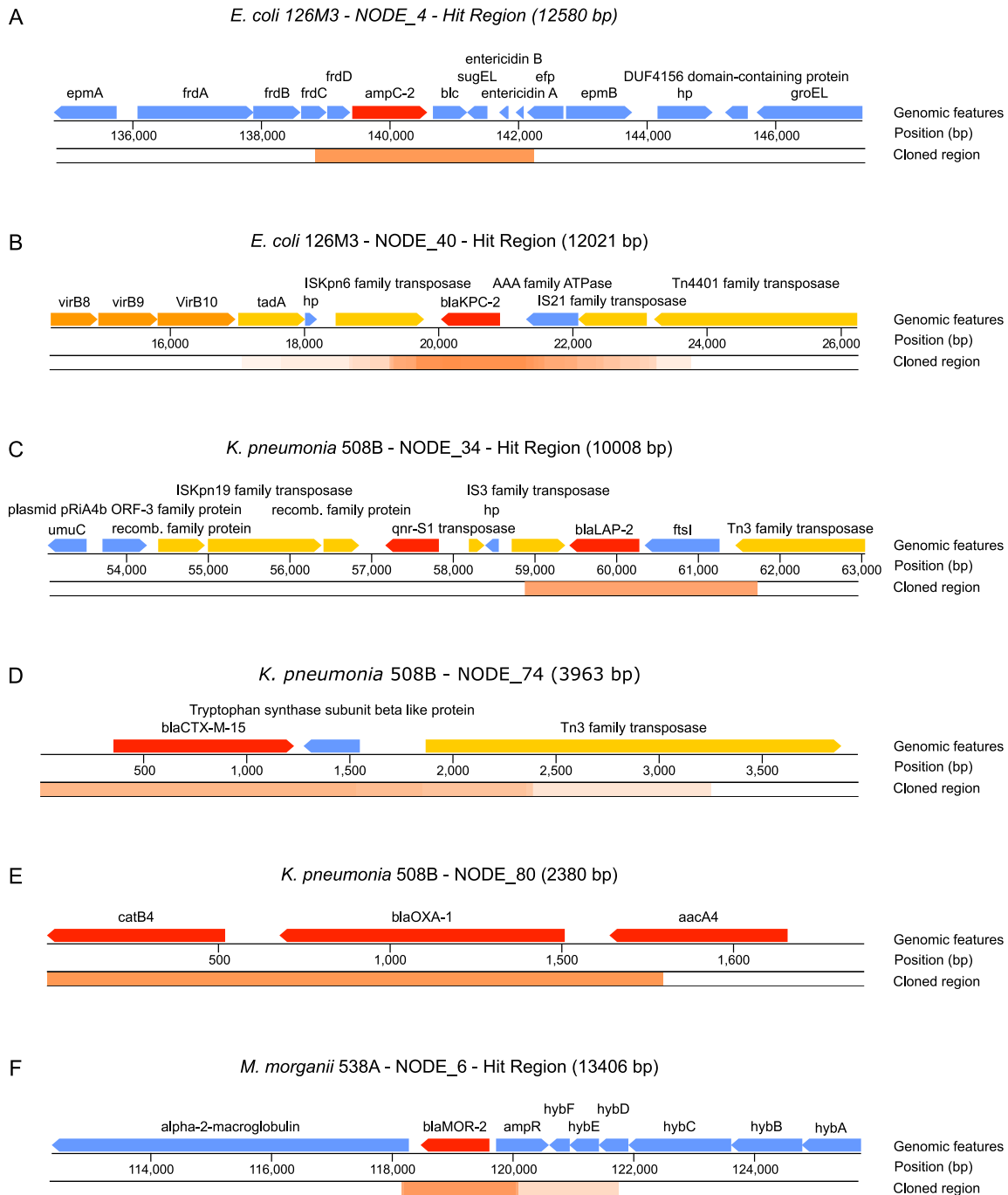


Figure 6. Features found in identified clones conferring resistance to antibiotics. **A)** *ampC-2* gene identified from *E. coli* 126M3. **B)** *bla_{KPC-2}* gene identified from *E. coli* 126M3. **C)** *bla_{LAP-2}* gene identified from pKP508BN34 plasmid from *K. pneumoniae* 508B. **D)** *bla_{CTX-M-15}* gene identified from *K. pneumoniae* 508B. **E)** *bla_{OXA-1}* gene identified from *K. pneumoniae* 508B. **F)** *bla_{MOR-2}* gene identified from *M. morganii* 538A. ORFs were colored according to its function: red – ORFs directly related to antibiotic resistance; orange – ORFs related to virulence; yellow – ORFs related to horizontal transfer of the ARG; blue – ORFs with no identified relation to pathogenesis. The cloned region represents contig regions that were identified in our screenings. Overlapping cloned regions are darker, while dim regions have fewer overlapping identified clones.

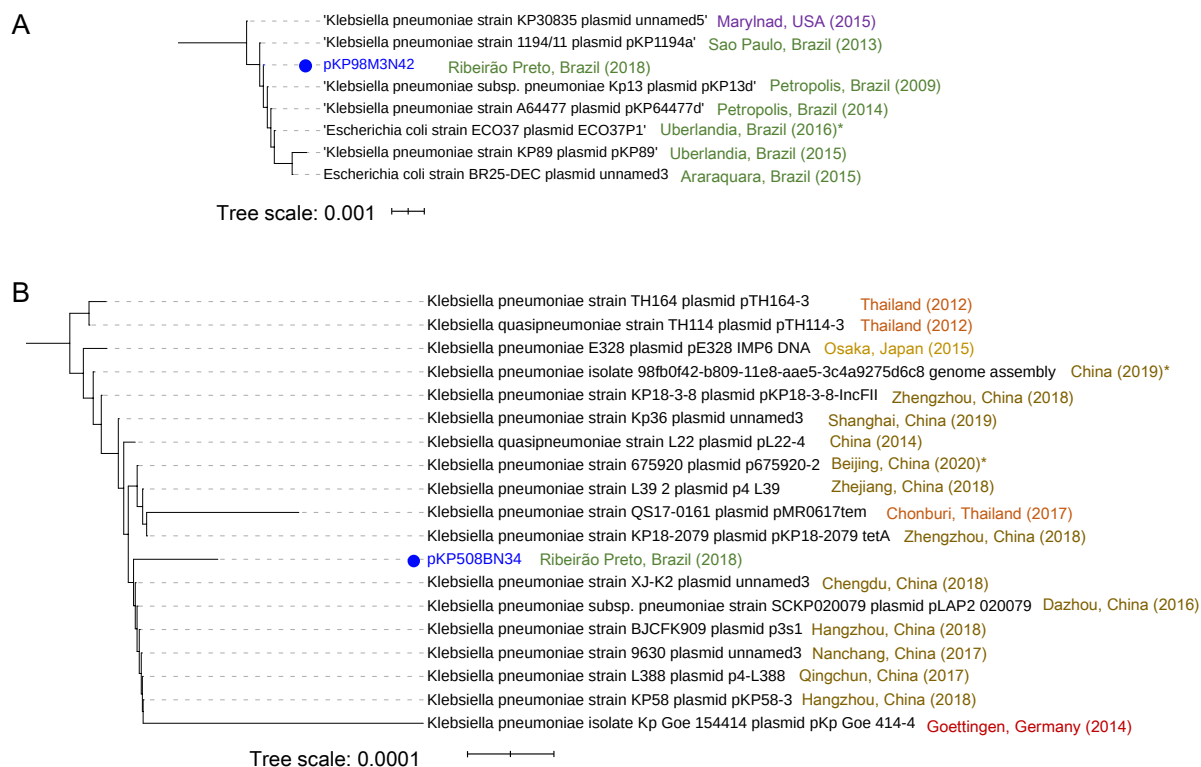
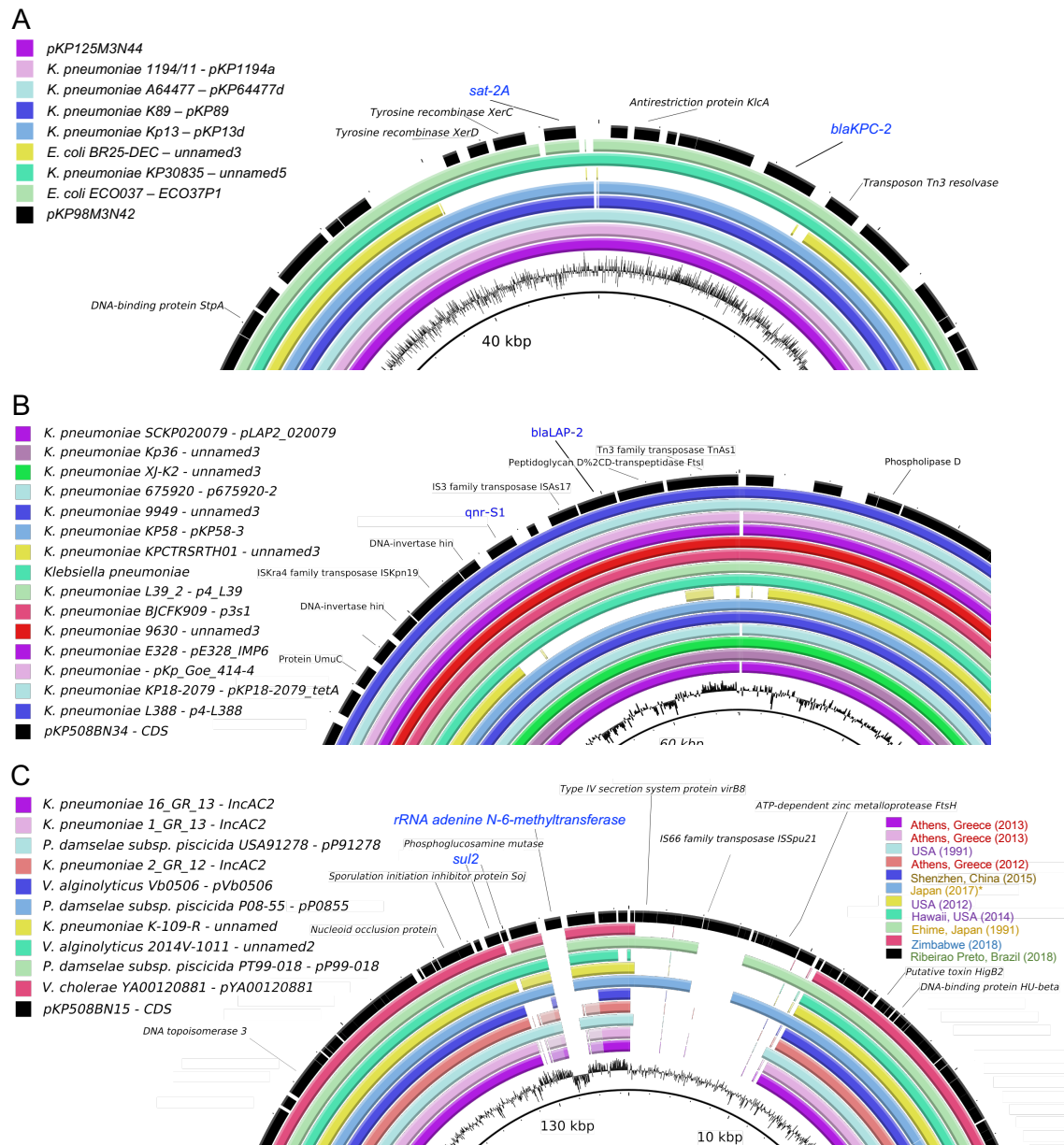


Figure 7. Distance relationship analysis for two plasmids from *K. pneumoniae*. **A)** Distance relationship of the 7 closest plasmids' nucleotide sequences to pKP98M3N42 showing an *E*-value of less than 0.0 and a minimal sequence cover of 70% in BLAST analysis. The tree was produced using pairwise alignments by means of the Fast-minimum evolution method. Year denotes the collection data and the asterisk indicates the data reported in the public database. **B)** Distance relationship of the 18 closest plasmids' nucleotide sequences to pKP508BN34 showing an *E*-value of less than 0.0 and a minimal sequence cover of 50% in BLAST analysis. The tree was produced using pairwise alignments by means of the Fast-minimum evolution method. Year states the collection data and the asterisk indicates the data reported in the public database, informed when collection data was not available. iTOL (<https://itol.embl.de>) was used for tree visualization.



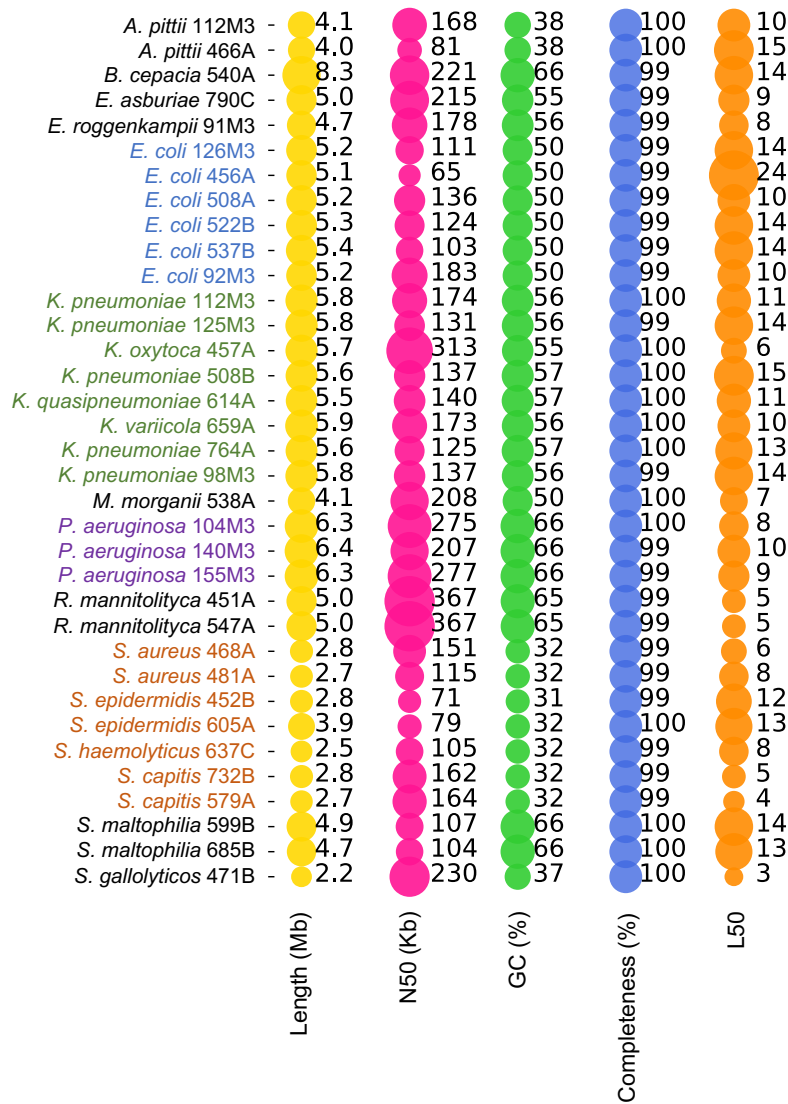


Figure S1. General features of genomes sequenced in this study.

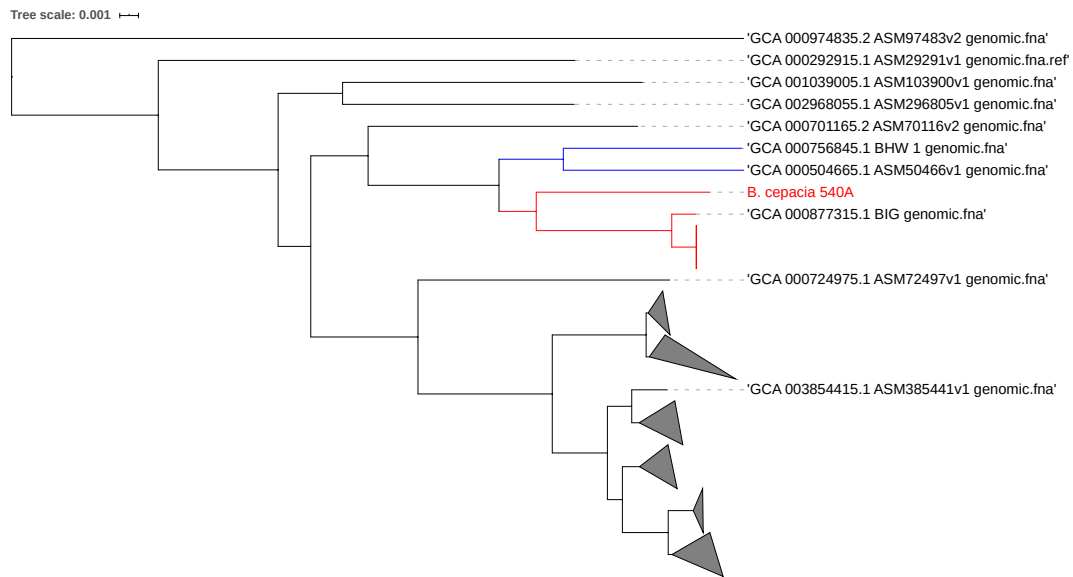


Figure S2. Phylogenetic tree of *B. cepacia* 540A and total number of available *B. cepacia* genomes (166) in NCBI at March 2020. For simplicity, branch with high similarity were compressed (gray triangles). NCBI accession number are indicated for *B. cepacia* strains compared in the figure. Blue branches are two *B. cepacia* endophytic strains isolated in Australia, while red branches are *B. cepacia* isolated from patients.

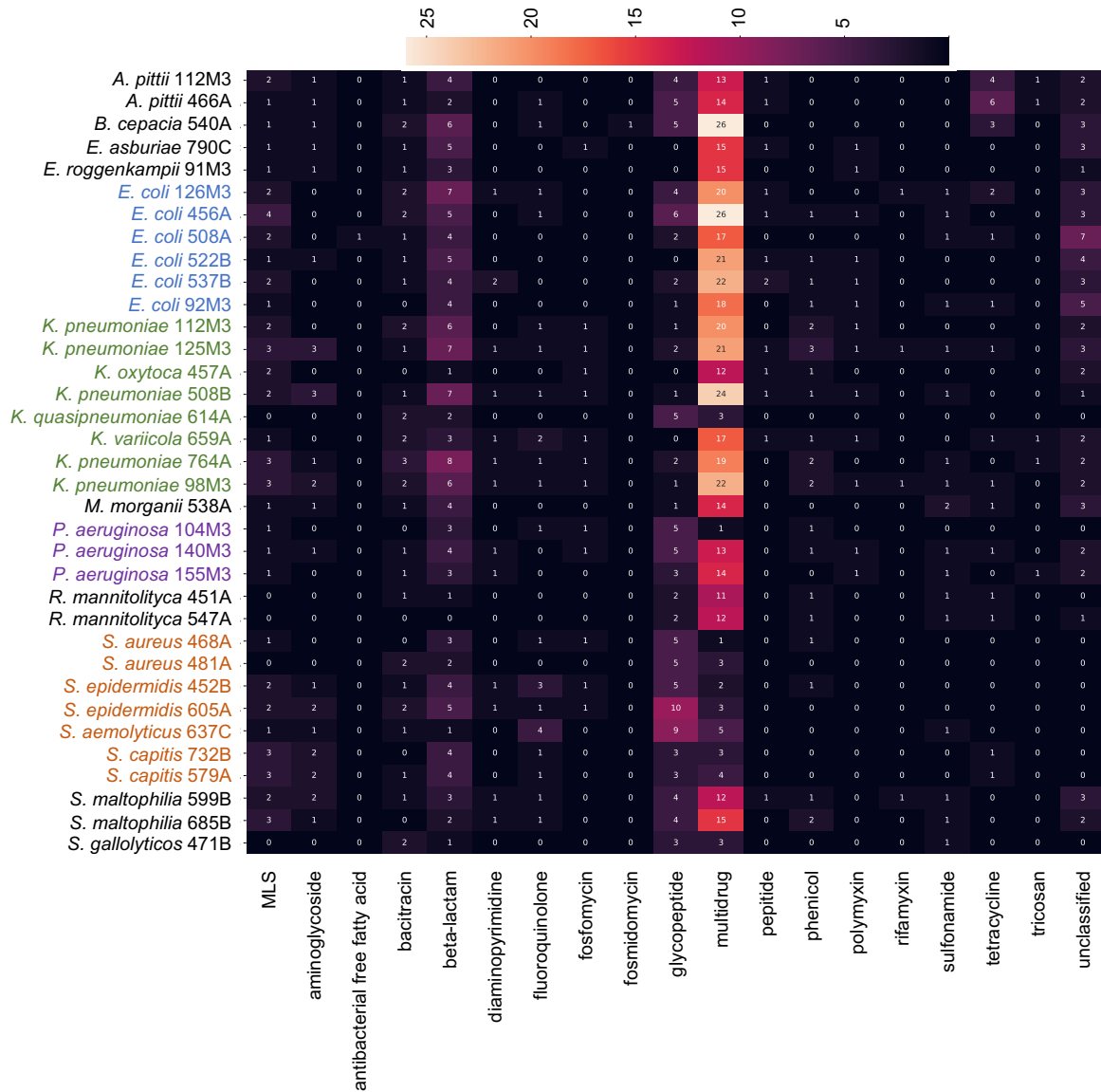


Figure S3. Heatmap showing the presence of ARGs identified by DeepARG in nearly sequenced genomes.

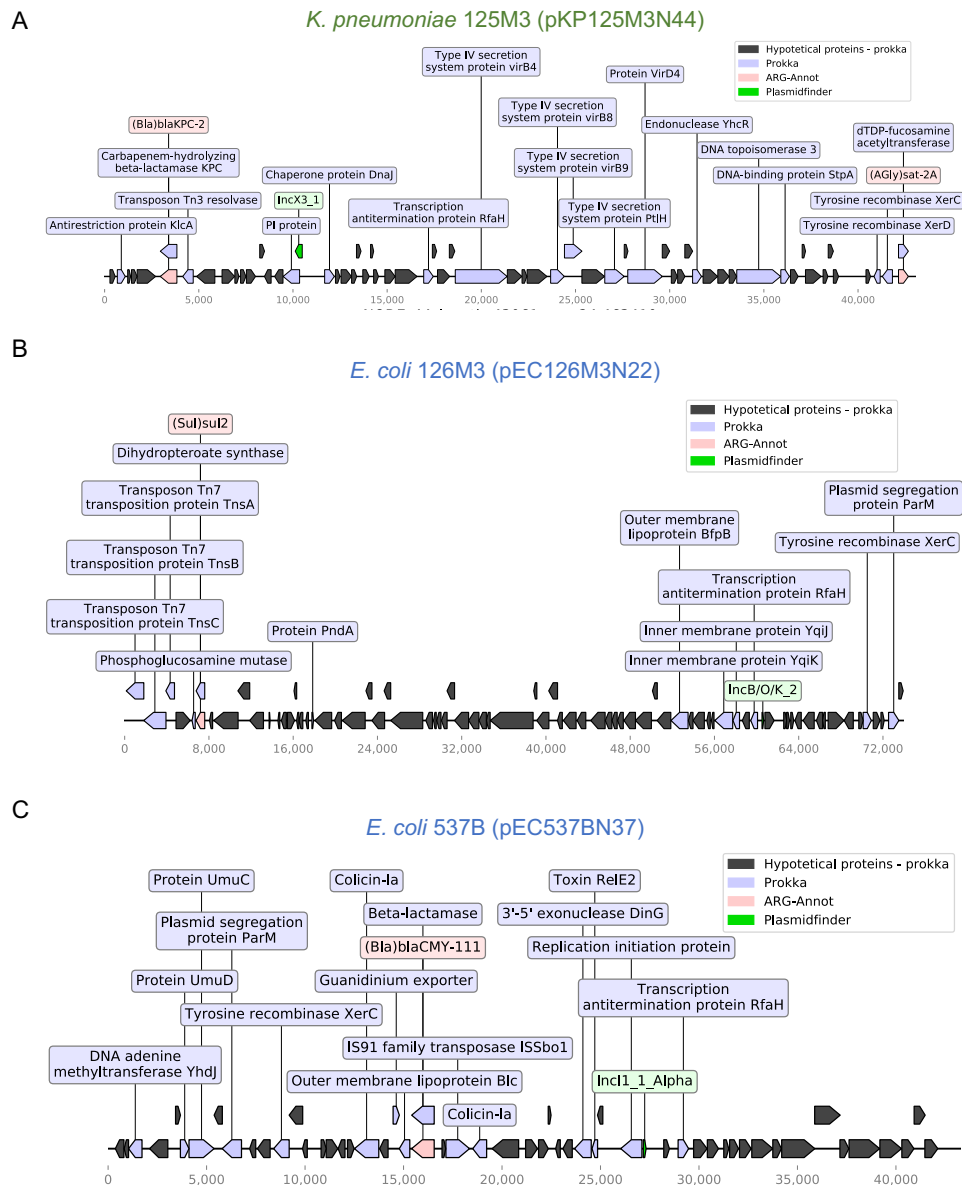


Figure S4. Schematic representation of genes in identified plasmids exposing the coexistence of ARGs with transposon elements. **A)** Plasmid pKP125M3N44 (43 kb) from *K. pneumoniae* 125M3. This plasmid is identical to plasmid pKP98M3N42 from *K. pneumoniae* 98M3 (Fig. 4A). **B)** Plasmid pEC126M3N22 (73,9 kb) from *E. coli* 126M3. **C)** Plasmid pEC537BN37 (43,3 kb) from *E. coli* 537B. Legends represents the colors code for the identified genes.

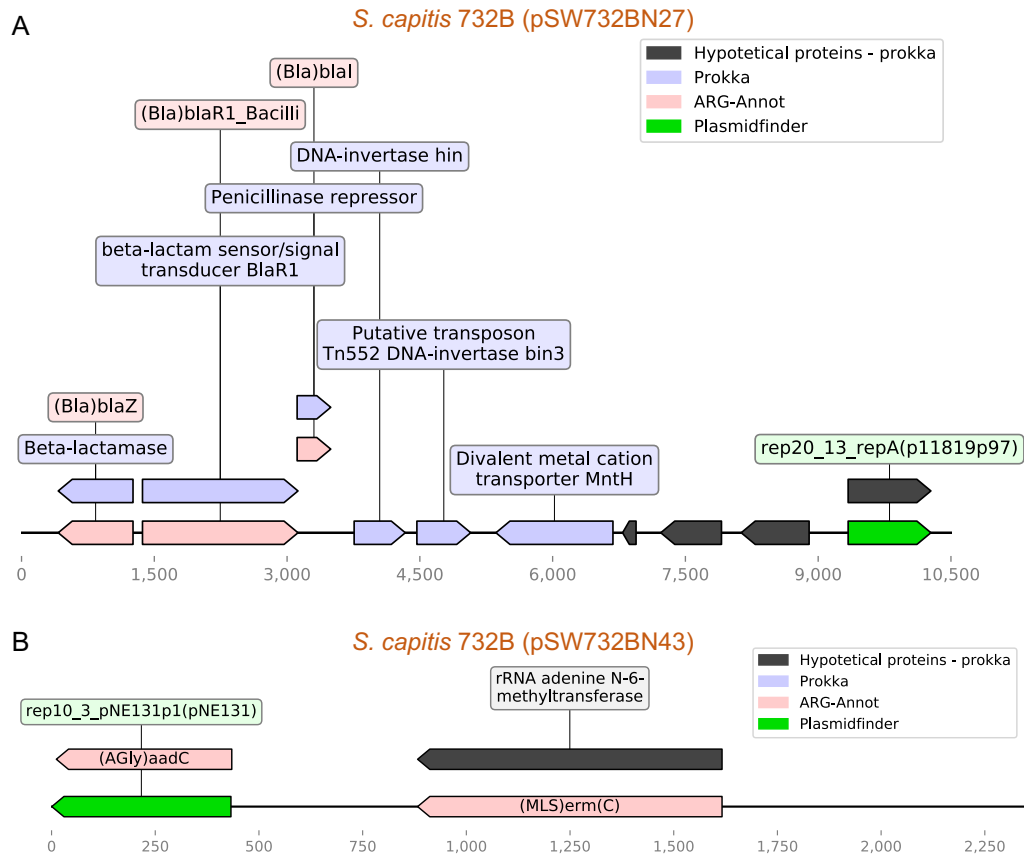


Figure S5. Schematic representation of plasmids identified in *Staphylococcus* strains showing the coexistence of ARGs with transposon elements. **A)** Plasmid pSW732BN27 (10.5 kb) from *S. capitis* 732B. **B)** Plasmid pSW732BN43 (~2.3kb) from *S. capitis* 732B.