



Data Article

Plasmid sequence dataset of multidrug-resistant Enterobacterales isolated from hospital effluents and wastewater treatment plant

Christopher Mutuku^a, Barbara Kutasy^b, Peter Urban^c, Szilvia Melegh^d, Robert Herczeg^c, Zoltan Gazdag^a, Eszter Virág^{e,*}

^a Department of General and Environmental Microbiology, Faculty of Sciences, University of Pécs, 7622 Pécs, Hungary

^b Department of Plant Physiology and Plant Ecology, Hungarian University of Agriculture and Life Sciences, Georgikon Campus Keszthely, 8360 Keszthely, Festetics Gy. u.7., Hungary

^c Bioinformatics Research Group, Szentágotthai Research Centre, 7624 Pécs, Hungary

^d Department of Medical Microbiology and Immunology, Medical School, University of Pécs, 7622, Pécs, Hungary

^e Department of Molecular Biotechnology and Microbiology, Institute of Biotechnology, Faculty of Science and Technology, University of Debrecen, Egyetem Square 1, 4032 Debrecen, Hungary

ARTICLE INFO

Article history:

Received 28 September 2022

Revised 2 November 2022

Accepted 4 November 2022

Available online 9 November 2022

Dataset link: [Enterobacterales plasmids - raw sequence reads \(Original data\)](#)

Keywords:

Hospital effluents

Wastewater treatment plant

Enterobacterales

β -lactamases

Resistance genes

Plasmid

ABSTRACT

We present plasmid sequences of 21 multidrug resistant isolates of Enterobacterales belonging to *Escherichia coli* (n=10), *Klebsiella pneumoniae* (n=9), *Klebsiella oxytoca* (n=1), and *Citrobacter freundii* (n=1). The isolates originated from effluent collected from hospital sewer pipes and from a wastewater treatment plant (WWTP) in a southwestern Hungarian city. Isolation was carried out using eosin methylene blue agar supplemented with ceftriaxone and the isolates were identified with MALDI-TOF MS. Screening for multidrug resistance was conducted by determining susceptibility to four chemical classes namely, beta-lactams, aminoglycoside, fluoroquinolone, and sulfonamide. Plasmid DNA was isolated by alkaline lysis method using the Monarch plasmid DNA miniprep kit from freshly grown pure colonies. Molecular

* Corresponding author.

E-mail address: virag.eszter@science.unideb.hu (E. Virág).

typing and Illumina sequencing of plasmid DNA of multiresistant strains were performed. After the assembly of contigs, genes localized on plasmid sequences were determined and functionally annotated. These reconstructed plasmid sequences supplemented with gene functional annotations were deposited in the Mendeley data. Using these datasets different plasmid incompatibility groups were identified. These conjugative plasmids appear to play a key role in the transmission of multiple resistance genes in enteric bacteria via wastewater. The presented data may provide useful insight on the correlations between environmental antibiotic contamination and the development of bacterial resistance, which poses a serious public health threat.

© 2022 The Author(s). Published by Elsevier Inc.

This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>)

Specifications Table

Subject	Microbiology: Applied Microbiology
Specific subject area	Wastewater from human sources may serve as an important reservoir of multiresistant Enterobacterales. The bacterial strains were cultured with an aim to isolate and sequence plasmids harboring antibiotic resistance genes.
Type of data	Tables Database record Figure
How the NGS data were acquired	Illumina sequencing of plasmid DNA of 21 bacterial strains was performed on the NovaSeq 6000 instrument (Illumina, San Diego, CA, USA) with a 2 × 151 run configuration. Quality control (QC), trimming, and filtering of 150 bp paired-end raw reads were performed by using FastQC and Trimmomatic software programs. For plasmid identification, genes characteristically encoded on plasmids for each strain were determined based on literature and by aligning them to the contigs using locally the Blast+ program. Prokaryotic gene finding was performed by Climmer and their functional annotation and Gene Ontology (GO) analysis were carried out using OmixBox.Biobam software package.
Data format	Raw Analyzed Filtered
Description of data collection	The multiresistant enteric bacterial isolates were collected from the effluent of hospital sewer pipes and from a wastewater treatment plant (WWTP) in a southwest Hungarian city. Illumina sequencing was performed for the plasmids isolated from the multiresistant strains. The reconstructed plasmid sequences of 21 resistant strains are reported here.
Data source location	<ul style="list-style-type: none"> • Institution: Department of General and Environmental Microbiology, Faculty of Sciences, University of Pécs. • City/Town/Region: Pécs, Baranya, 7622 • Country: Hungary
Data accessibility	<p>The BioProject and sequence reads are available in National Center for Biotechnology Information (NCBI) database under the accessions:</p> <p>Repository name: Enterobacterales plasmids – raw sequence reads Data identification number: PRJNA877861 Direct link to dataset: https://www.ncbi.nlm.nih.gov/sra/PRJNA877861 Repository name: Enterobacterales Plasmids, Mendeley Data, V1 Data identification number: doi:10.17632/j3mkwhzh84.1 Direct URL to data: https://data.mendeley.com/datasets/j3mkwhzh84/1</p>

(continued on next page)

Related research article

Instructions for accessing these data: The contigs of identified plasmid sequences of 21 samples can be found in the file: Plasmid_sequences (1).rar. Annotation of these contigs can be found in the file: Omicsbox_annot_table.zip
Christopher Mutuku, Szilvia Melegh, Krisztina Kovacs et al. Characterization of β -Lactamases and Multidrug Resistance Mechanisms in Enterobacterales from Hospital Effluents and Wastewater Treatment Plant (2022) *Antibiotics* 11, 776. 6.
<https://doi.org/10.3390/antibiotics11060776>

Value of the Data

- These data shed light on the current and developing trends in acquired antimicrobial resistance in clinically important bacteria including *E. coli*, *Klebsiella spp.*, and *Enterobacter spp.*, which are of human origin and enter the environment via wastewater. It can form a basis for comparison with clinical surveillance data, which is useful for epidemiological studies.
- The presence of plasmid mediated resistance among enteric bacteria in wastewater, which primarily form part of human and animal gut microbiota, is subject to further investigation to ascertain whether (1) the opportunistic pathogens isolated already carried these genes from the source population of the wastewater or (2) horizontal gene transfer in the environment lead to acquisition of resistance genes by these human associated microbes from the environmental microbes.
- Pharmaceutical antimicrobials in untreated hospital effluents whose presence is likely to create selection pressure leading to the emergence of multidrug resistant bacteria (ARB) demonstrated in our data require to be investigated to give an insight into the possible correlation between environmental antibiotic contamination and the development and spreading of resistance.
- The selection of isolates for next generation sequencing was based on the following considerations: (1) number of antimicrobials that the isolate was resistant to (higher number being preferred) and, (2) the frequency of isolation (higher occurrence being preferred). In this study, *E. coli* (n=58), *K. pneumoniae* (n=26), *K. oxytoca* (n=17) and *C. freundii* (n=14) were the most frequently identified species. The selected *E. coli* (n=10), *K. pneumoniae* (n=9), *K. oxytoca* (n=1) and *C. freundii* (n=1) isolates that demonstrated outstanding multiresistance, were considered to be the most suitable candidates for NGS as shown in [Table 1](#).

1. Data Description

The origin of the 21 resistant enteric bacterial strains is detailed in [Table 1](#). The reconstructed plasmid sequences of these samples are deposited to the Mendeley Data, Plasmid_sequences (1).rar. Prokaryotic genes presented on these plasmids with GO annotation are deposited on Mendeley Data, Omicsbox_annot_table.zip. Sequences in .fasta format and an annotation table in .xlsx format under the doi number: 10.17632/j3mkwhzh84.1.

Different plasmid incompatibility groups, including IncB/O/K/Z, IncFIA, IncFIB, and IncFII, carrying multiresistant genes were identified, suggesting a key role in the transmission of multiple resistance genes in enteric bacteria via wastewater. The isolates were clustered in accordance with the resistance genes harbored ([Fig. 1](#)).

Table 1
Isolates, collection date and sources of samples.

Serial number	Strain	Isolate	Sampling location	Collection date
1	CF102	<i>Citrobacter freundii</i>	DGSL	17-7-2019
2	EC10	<i>Escherichia coli</i>	H3	10-4-2019
3	EC11	<i>Escherichia coli</i>	H3	31-10-2019
4	EC14	<i>Escherichia coli</i>	H3	31-10-2019
5	EC19	<i>Escherichia coli</i>	ACSL	17-7-2019
6	EC20	<i>Escherichia coli</i>	ACSL	17-7-2019
7	EC66	<i>Escherichia coli</i>	H4	16-5-2019
8	EC74	<i>Escherichia coli</i>	NH	15-10-2019
9	EC75	<i>Escherichia coli</i>	NH	15-10-2019
10	EC81	<i>Escherichia coli</i>	H2	18-6-2019
11	EC92	<i>Escherichia coli</i>	H1	13-6-2019
12	KO54	<i>Klebsiella oxytoca</i>	H4	16-5-2019
13	KP2	<i>Klebsiella pneumoniae</i>	H1	13-6-2019
14	KP4	<i>Klebsiella pneumoniae</i>	INFL	12-11-2019
15	KP45	<i>Klebsiella pneumoniae</i>	H2	18-6-2019
16	KP57	<i>Klebsiella pneumoniae</i>	H4	28-10-2019
17	KP76	<i>Klebsiella pneumoniae</i>	H4	28-10-2019
18	KP79	<i>Klebsiella pneumoniae</i>	H4	28-10-2019
19	KP93	<i>Klebsiella pneumoniae</i>	H2	23-10-2019
20	KP96	<i>Klebsiella pneumoniae</i>	H2	23-10-2019
21	KP6	<i>Klebsiella pneumoniae</i>	H1	8-10-2019

H1-H4; hospital effluents, NH: nursing home, INFL: influent, ACSL: activated sludge, DGSL: digested sludge

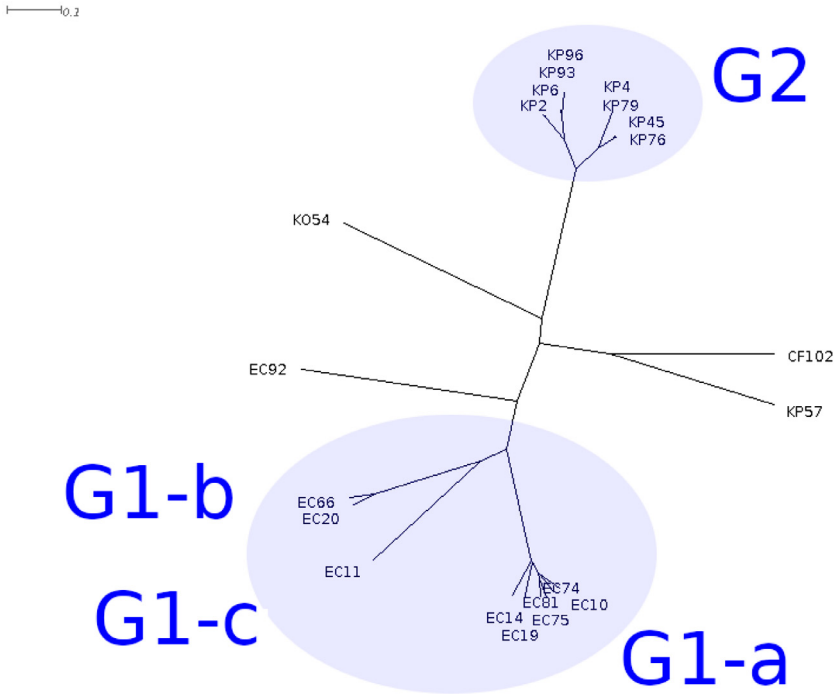


Fig. 1. Representation of clusters based on MUMi distance of plasmid sequences.

2. Wastewater Treatment Plant Operational Parameters

The WWTP located in a southwest Hungarian city serves a population equivalent to slightly over 200,000 inhabitants. The wastewater treatment involves three stages (primary clarification, secondary-activated sludge system, and tertiary-UV treatment), with the final effluent discharged into the nearby surface stream. The wastewater quality data is shown in Tables 2 and 3. In addition, anaerobic digestion occurs in large tanks designed to operate with continuous input of untreated sludge and removal of the final, stabilized sludge product. The digesters operate in the mesophilic range, with the regulated heating kept at 36.5 - 37.5°C. The technology is co-digestion fermentation, i.e. in addition to the sludge produced at the sewage treatment plant, other organic wastes with a high organic matter content that are easily biodegradable for biogas production are used. Typical wastes treated at the plant include: sludge from other sewage plants, fatty flotation from the dairy industry, whey, kitchen and restaurant waste (swill), sludge from grease trap, slaughterhouse waste and sludge from sewage treatment plants in slaughterhouses.

At the initial step, the raw (primary sludge settled in the pre-settler) and excess sludge (secondary sludge produced during biological treatment) generated at the sewage plant are mixed and gravity-compressed in a 1500 m³ dorr-type stick sedimentation basin. The raw sludge with a dry matter content of 1.5% is removed from the pre-settler, as well as the redundant sludge with a dry matter content of 0.85%. After their mixing and gravity compression, a mixed sludge with an average dry matter content of 2.6% is obtained. In the biogas plant, this sludge is further thickened to a dry matter content of 5.5 - 6 % with the addition of polyelectrolyte using pre-dewatering tables. The pre-dewatered sludge goes into a 70 m³ homogenizing tank, where it is mixed with the received crushed waste. The waste is pre-treated with a specially developed device. The technology has a 30 m³ receiving container where the waste is emptied. A hardox plate with 8 mm perforation is placed at the bottom of the receiving bunker, above which two grinding wheel like devices with a diameter of 1 m rotate. The waste is shredded and pulped with this equipment, the required 5.5-6% dry matter content is adjusted by diluting with water and the resultant mixture flowing through the perforation is pumped into the homogenizing

Table 2

Water quality data on physico-chemical parameters during the sampling period.

	Chemical oxygen demand			Biochemical oxygen demand			Ammonium-nitrogen			Kjeldahl nitrogen		
	mg/l			mg/l			mg/l			mg/l		
	1	2	3	1	2	3	1	2	3	1	2	3
Mean	928.8	534.5	28.4	483.5	325.9	5.2	60.45	64.5	0.53	82.6	79.4	2.9
	Nitrite mg/l		Nitrate mg/l	Total phosphorus mg/l			pH mg/l			Total suspended solids mg/l		
	3	3	1	2	3	1	2	3	1	2	3	
Mean	0.58	39.59	9.6	8.6	0.44	7.74	7.77	7.62	366.8	139.9	4.57	

1 = incoming raw sewage sample, 2 = sample taken after mechanical cleaning, before biological cleaning, 3 = treated wastewater sample.

Table 3

Other WWTP operational parameters

Average amount of incoming wastewater	24.600 m ³ /d.
Average wastewater treatment plant total residence time	17 hours
Biological reactor residence time	4.5 hours
Sludge age (summer - winter)	15 - 19 days
The proportion of recirculation sludge at the biological cleaning stage	1.5 times the incoming wastewater (the recirculation circle is 150%).

tank. The homogenized sewage sludge and waste mixture is pumped into the 2 piece 4000 m³ mesophilic fermentation towers.

SRT: not interpretable in these mesophilic digesters.

HRT: varies between 19 and 24 days, depending on the amount of externally delivered waste.

3. Materials and Methods

3.1. Sampling strategy

Wastewater samples were drawn from four hospital wastewater discharge points, H1 (387 beds), H2 (106 beds), H3 (127 beds), and H4 (348 beds), discharge points of a nursing home for the elderly (NH, 490 beds), and from a wastewater treatment plant (WWTP). Effluent samples from the healthcare facilities were collected directly from two or three separate generation points serving different buildings before joining the main sewer pipe. Samples from the WWTP were collected from the influent directly behind the grating screen. A 30 mL sample was collected every 15 minutes by lowering a flask into the wastewater flow over a period of 4 h and the aliquots pooled to constitute 480 mL composite sample in sterile 500 mL glass bottles. One grab sample was drawn from the activated sludge reactor and the final product of the mesophilic digested sludge. The study was aimed at comparing the change in the presence of antibiotic resistance genes during wastewater treatment. Therefore, the samples were drawn from influent wastewater, the activated sludge reactor, and mesophilic digested sludge because the activated sludge reactor harbors the bacterial cells for hours, and they last for days in mesophilic fermentation towers. In both spaces, the metabolism of microbes is most intense, and there is a chance for gene exchange. Samples were transported on ice to the laboratory and stored at 4°C before assaying within 6 hours. The study was conducted in 2019.

3.2. Isolation of multiresistant *Enterobacterales*

Isolation was carried out using eosin methylene blue agar supplemented with ceftriaxone (2.0 µg ml⁻¹) [1] to select for the lactose fermenting colonies characteristic of the fecal coliform bacteria, including *E. coli*, *Klebsiella pneumoniae*, *Enterobacter* spp., and *Citrobacter* spp. The isolates were identified with MALDI-TOF MS (MALDI Biotyper, Bruker) according to the instructions by the manufacturer. Score value ≥2.0 was considered as species level identification. If the score value was less than 2.0, the measurement was repeated until species level identification was achieved. Screening for multidrug resistance was based on disk diffusion testing with four chemical classes namely, beta-lactams (ceftriaxone, CRO, 30 µg, ceftazidime, CAZ, 10 µg, cefotaxime, CTX, 30 µg, cefpodoxime, CPD, 10 µg, ceftazidime, FOX, 30 µg, imipenem, IMP, 10 µg, and meropenem, MEM, 10 µg), aminoglycoside (gentamicin, GN, 10 µg), fluoroquinolone (ciprofloxacin, CIP, 5 µg), and sulfonamide (sulfamethoxazole/trimethoprim, SXT, 1.25/23.75 µg) Oxoid, Wesel, Germany). Quality control was performed using known characterized strains. Source of the isolates selected for NGS are presented in [Table 1](#).

3.3. Plasmid DNA isolation

Fresh pure colonies grown on Mueller Hinton agar were transferred into Luria Bertani broth and incubated in an orbital shaker at 35 °C and 200 rpm for 12–16 h. All the centrifugation steps were carried out at 16,000 × g. Plasmid DNA was isolated using the Monarch plasmid DNA miniprep kit according to the manufacturer's instructions (New England Biolabs T1010, Ipswich, Massachusetts, USA). DNA concentration and purity were determined using a Nanodrop spectrophotometer (NanoDrop 2000, Thermo Scientific, Wilmington, NC, USA), and stored at -20 °C.

3.4. Sequencing library preparation

The sequencing library preparation of selected isolates (Table 1) was prepared using Swift 2S Turbo DNA Library Kits (Swift Biosciences, Ann Arbor, Michigan, United States). 100 ng genomic DNA was fragmented, end prepped, and adapter ligated. Magnetic bead size selection was performed to select 250–300 bp insert size fragments, followed by the library amplification according to the manufacturer's instructions. The quality of the library was checked on the 4200 TapeStation System using D1000 Screen Tape (Agilent Technologies, Palo Alto, CA, USA) and the quantity was measured on Qubit 3.0. (Thermo Scientific, Waltham, MA, USA). Illumina sequencing was performed on the NovaSeq 6000 instrument (Illumina, San Diego, CA, USA) with a 2×151 run configuration.

3.5. Pre-processing and assembly of reads

The QC analysis was performed using FastQC software (v0.11.9) [2]. Phred-like quality scores (Qscores) were set to >30 . Poor quality reads, adapters at the ends of reads, limited skewing at the ends of reads were eliminated by using Trimmomatic [3]. De novo assembly of cleaned reads was performed with SPAdes (v3.15.3) [4]. Since the data contained genomic DNA debris, identification of plasmid-derived contigs was performed after this step as described in Mutuku et al., 2022 [5]. Number, total length, and N50 of assembled contigs ranged between 17 and 64, 182 477–547 810 bp, and 11 403–62 343 bp, respectively.

3.6. Gene prediction

Prokaryotic gene identification was performed using Glimmer (v3.0) based on an interpolated Markov model [6]. Functional annotation and Gene Ontology (GO) analysis of identified genes were performed using OmixBox.Biobam [7] as described by Mutuku et al., 2022 in detail [5]. For detection of antimicrobial resistance genes and the identification of plasmid incompatibility groups, ResFinder (v4.1) and PlasmidFinder (v2.1) were used [8,9].

3.7. Clustering of isolates

MUM index (MUMi) was calculated as described by Marc Deloger et al. [10] Briefly nucmer (mummer-4.0.0) was used to search for all matches between plasmid sequences pairwise (dataset1 and dataset2) with the '-mum' option. MUMs were filtered to remove overlapping sequences. First MUMs completely included in larger MUMs were excluded; next MUMs completely included in adjacent MUMs were removed; at last MUMs with partial overlaps were trimmed for both plasmid sequence datasets by shifting the corresponding coordinates and recalculating MUM length in both datasets. The three-step filtering was done for each of the plasmid sequence pairs (first for dataset1 and then for dataset2). MUM indices were calculated as: $MUMi = 1 - L_{mum} / L_{av}$ where L_{mum} is the total of the length of all nonoverlapping MUMs in the plasmid sequence dataset and L_{av} stands for the average length for the two plasmid sequence datasets to be compared. The MUMi distance matrix was imported into SplitsTree4 and neighbor joining method was used for clustering of plasmid sequences.

According to MUM indices, the isolates were clustered into six groups, designated as G1–G6. The two main groups contained the majority of *E. coli* (G1) and were shown to harbor *bla*_{CTX-M-27} type ESBL gene, aminoglycoside (*aadA5*, *aph(3'')-Ib*, *aph(6)-Id*), folate inhibitor (*dfrA17*, *sul1*, *sul2*), tetracycline (*tet(A)*), macrolide (*mph(A)*), and quaternary ammonium compound (*qacEΔ*) resistance genes. Cluster G2 enclosed eight, closely related *K. pneumoniae* isolates. The presence of multiple β -lactamase genes (Extended spectrum β -lactamase-ESBL: *bla*_{CTX-M-27}, non-

ESBL: *bla*_{TEM-1}, *bla*_{OXA-1}), aminoglycoside (*aac*(6′)-*Ib-cr*, *aph*(3′′)-*Ib*, and *aph*(6)-*Id*), chloramphenicol (*catB3*), folate inhibitor (*dfrA14*, *sul2*), and quinolone (*qnrB1*, *aac*(6′)-*Ib-cr*) resistance genes. Four unrelated isolates, namely *C. freundii*, *K. pneumoniae*, *K. oxytoca*, and *E. coli*, were shown to harbor either *bla*_{CTX-M-15}, *bla*_{SHV-12}, *bla*_{CTX-M-30}, or *bla*_{CTX-M-1} ESBL genes, respectively. The clustering is visualized in [Fig. 1](#).

4. Limitation of the Study

Due to limited resources, a small number of multiresistant isolates were selected for NGS to provide representative information.

Ethics Statements

Not relevant for the data.

Declaration of Competing Interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data Availability

[Enterobacterales plasmids - raw sequence reads \(Original data\)](#) (Mendeley Data).

CRediT Author Statement

Christopher Mutuku: Investigation, Data curation, Methodology, Writing – original draft; **Barbara Kutasy:** Validation; **Peter Urban:** Methodology; **Szilvia Melegh:** Writing – review & editing, Data curation; **Robert Herczeg:** Software; **Zoltan Gazdag:** Conceptualization, Methodology, Resources, Writing – review & editing; **Eszter Virág:** Software, Writing – review & editing.

Acknowledgments

The research was performed in collaboration with Bioinformatics Research Group, Genomics and Bioinformatics Core Facility at the Szentágotthai Research Centre of the University of Pécs. We acknowledge all the personnel who guided the collection of wastewater samples including László Bicsák, Viktória Fébó in the wastewater treatment plant, Ferenc Illés, Gyula Tóth, József Zsámboki, Tamás Hrakovszky in the hospitals and the nursing home. Further, we acknowledge Csabane Boros for technical laboratory assistance.

References

- [1] C.G. Giske, J. Turnidge, R. Cantón, G. Kahlmeter, Update from the European committee on antimicrobial susceptibility testing (EUCAST), *J. Clin. Microbiol.* 60 (3) (2021) e00276-21.
- [2] S. Andrews, F. Krueger, A. Segonds-Pichon, L. Biggins, C. Krueger, S. Wingett, FastQC, 2010.
- [3] A.M. Bolger, M. Lohse, B. Usadel, Trimmomatic: a flexible trimmer for Illumina sequence data, *Bioinformatics* 30 (15) (2014) 2114–2120.
- [4] A.D. Prjibelski, I. Vasilinets, A. Bankevich, A. Gurevich, T. Krivosheeva, S. Nurk, S. Pham, A. Korobeynikov, A. Lapidus, P.A. Pevzner, ExSPAnDer: a universal repeat resolver for DNA fragment assembly, *Bioinformatics* 30 (12) (2014) i293–i301.

- [5] C. Mutuku, S. Melegh, K. Kovacs, P. Urban, E. Virág, R. Heninger, R. Herczeg, Á. Sonnevend, A. Gyenesei, C. Fekete, Z. Gazdag, Characterization of β -lactamases and multidrug resistance mechanisms in enterobacterales from hospital effluents and wastewater treatment plant, *Antibiotics* 11 (6) (2022) 776.
- [6] S.L. Salzberg, A.L. Delcher, S. Kasif, O. White, Microbial gene identification using interpolated Markov models, *Nucl. Acid. Res.* 26 (2) (1998) 544–548.
- [7] B. Bioinformatics, S. Valencia, OmicsBox-Bioinformatics made easy, *March* 3 (2019) 2019.
- [8] A. Carattoli, E. Zankari, A. García-Fernández, M. Voldby Larsen, O. Lund, L. Villa, F. Møller Aarestrup, H. Hasman, In silico detection and typing of plasmids using PlasmidFinder and plasmid multilocus sequence typing, *Antimicrob. Agent. Chemother.* 58 (7) (2014) 3895–3903.
- [9] V. Bortolaia, R.S. Kaas, E. Ruppe, M.C. Roberts, S. Schwarz, V. Cattoir, A. Philippon, R.L. Allesoe, A.R. Rebelo, A.F. Florensa, ResFinder 4.0 for predictions of phenotypes from genotypes, *J. Antimicrob. Chemother.* 75 (12) (2020) 3491–3500.
- [10] M. Deloger, M.E. Karoui, M-A Petit, A genomic distance based on MUM indicates discontinuity between most bacterial species and genera, *J. Bacteriol.* 191 (1) (2009) 91–99.