Article

Plasmid Identification and Plasmid-Mediated Antimicrobial Gene Detection in Norwegian Isolates

Abdolrahman Khezri [1] (ID)**, Ekaterina Avershina** [1,2] (ID) **and Rafi Ahmad** [1,3,4,*] (ID)

1   Department of Biotechnology, Inland Norway University of Applied Sciences, 2318 Hamar, Norway;
    Abdolrahman.khezri@inn.no (A.K.); ekaterina.avershina@inn.no (E.A.)
2   Laboratory or Postgenomic Technologies, Izmerov Research Institute of Occupational Health,
    105275 Moscow, Russia
3   Division of Medical Services-Clinical Microbiology, Inland Hospital, 2609 Lillehammer, Norway
4   Institute of Clinical Medicine, Faculty of Health Sciences, UiT—The Arctic University of Norway, Hansine
    Hansens veg 18, 9019 Tromsø, Norway
*   Correspondence: Rafi.Ahmad@inn.no; Tel.: +47-6251-7845

Abstract: Norway is known for being one of the countries with the lowest levels of antimicrobial resistance (AMR). AMR, through acquired genes located on transposons or conjugative plasmids, is the horizontal transmission of genes required for a given bacteria to withstand antibiotics. In this work, bioinformatic analysis of whole-genome sequences and hybrid assembled data from *Escherichia coli,* and *Klebsiella pneumoniae* isolates from Norwegian patients was performed. For detection of putative plasmids in isolates, the plasmid assembly mode in SPAdes was used, followed by annotation of resulting contigs using PlasmidFinder and two curated plasmid databases (Brooks and PLSDB). Furthermore, ResFinder and Comprehensive Antibiotic Resistance Database (CARD) were used for the identification of antibiotic resistance genes (ARGs). The IncFIB plasmid was detected as the most prevalent plasmid in both *E. coli,* and *K. pneumoniae* isolates. Furthermore, ARGs such as *aph(3″)-Ib*, *aph(6)-Id*, *sul1*, *sul2*, *tet(D)*, and *qnrS1* were identified as the most abundant plasmid-mediated ARGs in Norwegian *E. coli* and *K. pneumoniae* isolates, respectively. Using hybrid assembly, we were able to locate plasmids and predict ARGs more confidently. In conclusion, plasmid identification and ARG detection using whole-genome sequencing data are heavily dependent on the database of choice; therefore, it is best to use several tools and/or hybrid assembly for obtaining reliable identification results.

Keywords: *Escherichia coli*; *Klebsiella pneumoniae*; antimicrobial resistance; plasmid; plasmid-mediated genes; horizontal gene transfer

1. Introduction

Antimicrobial resistance (AMR) is the ability of microorganisms to resist antimicrobial treatments, especially antibiotics. Infections due to AMR bacteria are a threat to modern health care and are responsible for an estimated 700,000 and 33,000 deaths/year globally and in Europe, respectively [1]. Recently, the World Health Organization (WHO) published a list of pathogens for which urgent global action is needed [2]. Extended spectrum β-lactamases (ESBL) producing and carbapenem-resistant *Enterobacteriaceae* are among the priority one critical section of the WHO pathogen list. There has been a global rise in infections caused by multi-drug resistant clones of *Enterobacteriaceae*, particularly *Klebsiella pneumoniae* and *Escherichia coli* [3].

AMR can arise through various mechanisms, including mutations of chromosomal genes and the acquisition of antibiotic resistance genes (ARGs) from other strains in a process termed horizontal gene transfer (HGT). It is the sharing of genes through HGT that has largely contributed to the global dissemination of ARGs [4]. The genomes of *E. coli* and *K. pneumoniae* are prone to a mutation in stress, depicting the genetic flexibility to upregulate

their natural resistance and acquire foreign determinants through HGT due to mobile genetic elements. These elements, such as plasmids, transposons, integrons, and genomic islands, harbor ARGs [5]. Several plasmids like IncF and IncI1 plasmids are known to carry resistance genes in *E. coli*, *K. pneumoniae,* and other *Enterobacteriaceae* [6]. Additionally, the ColE plasmids, which encode colicins, and have killing activity against different bacteria, are also important plasmids [7]. Broad range resistance plasmids are known to be associated with pathogens; for example, a resistance plasmid from *Enterobacteriaceae* can be transferred to a wide variety of Gram-negative organisms.

Whole-genome sequencing (WGS) is an effective method of tracking the onward transmission of bacteria or resistance plasmid transfer between bacteria. It has made it possible to determine and evaluate an organism's whole DNA sequence at low costs in a short period of time. It allows for the identification of antimicrobial resistance and the early detection of outbreaks or their epidemiological investigation [8]. Moreover, plasmid assembly and characterization following WGS is a difficult task. This happens because the plasmids tend to contain repeat sequences with sizes greater than sequences generated by sequencing platforms such as Illumina technology (San Diego, CA, USA) [9]. Therefore, the need for in silico plasmid detection has emerged due to the difficulty of plasmid DNA purification if they are longer than 50 kbp [10]. In addition, regarding the need for an efficient plasmid identification tool, ARG databases with comprehensive and accurate gene records are needed to assess AMR prevalence. Although several ARG databases exist, Comprehensive Antibiotic Resistance Database (CARD), and ResFinder are the most effective and have sustainable curation strategies [11]. Recent studies have shown that the hybrid assemblies, which are a combination of Illumina and long-read sequencing (e.g., Oxford Nanopore Technology's MinION) data, are better at identifying plasmids and ARGs [12]. However, this requires advanced bioinformatic and machine learning methods for WGS data analysis [13–15].

Globally, AMR is unevenly distributed. Recently, Klein et al. investigated the drug resistance index (DRI) for 41 countries [16]. They have compared the reported data on antibiotics' use and their resistance to the treatment of infections caused by microorganisms from the WHO priority list [2]. Norway is among the countries with the lowest DRIs (third lowest), and has a DRI value around four-fold lower than that of the country with the highest DRI; India. However, there is an increasing trend in AMR cases in Norway. For example, the percentage of *E. coli* with ESBL, causing septicemia, has increased ten-fold in the last ten years [17]. There has also been a slightly increased prevalence of ESBLs for *E. coli* (6.6% in 2017 and 6.5% in 2018) and *Klebsiella spp.* (5.3% in 2017 and 6.6% in 2018) [17].

This research has utilized different tools and databases to identify plasmids and predict plasmid-mediated ARGs in both *E. coli* and *K. pneumoniae* isolates. Our results indicate that plasmid identification and ARGs prediction are database/tool dependent. In this regard, a hybrid assembly can be considered an efficient way to identify plasmids and predict plasmid-mediated ARGs.

2. Materials and Methods

2.1. Sample Collection and Characterization

In this study, *E. coli* and *K. pneumoniae* isolates were collected from blood and urine specimens of Norwegian patients, in collaboration with Oslo University Hospital. The sample overview is in Table 1.

Table 1. An overview of the number of clinical isolates collected from Norwegian patients.

Isolates	Source	
	Blood	**Urine**
E. coli	53	5
K. pneumoniae	40	4

2.2. Library Preparation and Whole-Genome Sequencing

The WGS data used in this study are from our recent work, which was performed at Oslo University Hospital [18]. In brief, DNA was isolated from bacteria colonies using QIAamp DNA minikit (Qiagen, Hilden, Germany) following the manufacturer's instructions and was quantified using Qubit fluorometer (Life Technologies, Carlsbad, CA, USA). The libraries were constructed using the Nextera XT kit (Illumina Ltd., San Diego, CA, USA) according to the manufacturer's recommendations. The libraries were sequenced in pair-end mode (2×300 bp) using the Illumina MiSeq platform at the Norwegian Sequencing Center (Oslo, Norway). Furthermore, to make a hybrid assembly, we sequenced three more isolates (*E. coli* 39, *K. pneumoniae* 23, and 27) using the nanopore and the Illumina MiSeq sequencing platforms. Details regarding library preparation, sequencing, and hybrid assembly have previously been reported [12]. All bioinformatic analyses for both plasmid and hybrid assemblies were identical and performed as described below.

2.3. Bioinformatic Analyses of Bacterial Genomic

2.3.1. Quality Control and Trimming of Illumina Sequences

Initially, Illumina and Nanopore reads were quality checked using FastQC (v 0.11.8 for Linux) [19]. Then Illumina adapters were removed, and low-quality reads (Phred below 25) were filtered out using Trimmomatic with default parameters [20]. Before downstream analyses, trimmed reads were again quality checked using FastQC software.

2.3.2. Plasmid Assembly and in Silico Plasmid Identification

Putative plasmid sequences were assembled using plasmid flag in SPAdes (v 3.14.1 for Linux) [21]. General statistics of the assembled putative plasmids was assessed using QUAST (v 4.6.0 for Linux) [22]. Putative plasmid sequences were further confirmed using PlasmidFinder (software version: 2.0.1, database version: 2020-07-13) with minimum identity and coverage of 95% and 60%, respectively [23]. In addition to PlasmidFinder, the identification of putative plasmids was performed using two other methods. First, plasmid reference sequences were downloaded from the PLSDB database [24] and a curated database developed by Brooks et al. [25], hereafter referred to as Brooks. Later, assembled putative plasmids were BLAST searched (sequence identity >95% and word size 28) against downloaded reference plasmids databases. Initially, hits (contigs) with coverage between 30 to 100% were extracted and utilized for the next step. Then only hits with qcov $\geq$90% (P_{TRUE}) were considered for downstream analysis. The qcov is unique query coverage per subject, calculated after considering any alignment overlaps between different fragments aligned with that specific subject in the database.

2.3.3. Identification of Plasmids Mediated Antimicrobial Resistance Genes (ARGs)

To identify ARGs hosted by plasmids, the assembled putative plasmids for each isolate were submitted to Resfinder 4.0 [26] and resistance gene identifier tool from Comprehensive Antibiotic Resistance Database—CARD [27]. In both Resfinder and CARD, only hits showing $\geq$95% identity and $\geq$98% length coverage were considered as ARGs. Later, only hits sharing the same contig with P_{TRUE} were regarded as true plasmid-mediated ARGs (ARG_{Plasmid}). Hereafter we refer to $ARG_{\text{Plasmid-PlasmidFinder}}$ (meaning ARG and P_{TRUE} from PlasmidFinder were found on the same contig), $ARG_{\text{Plasmid-Brooks}}$ (meaning ARG and P_{TRUE} from Brooks were found on the same contig), $ARG_{\text{Plasmid-PLSDB}}$ (meaning that ARG and P_{TRUE} from PLSDB were found on the same contig).

3. Results

In the current study, putative plasmid sequences and ARGs were identified in silico in *E. coli* and *K. pneumoniae* isolates from Norwegian patients. Furthermore, hybrid assemblies from additional three isolates were also analyzed.

3.1. General Statistics of Assembled Plasmid Sequences and Hybrid Assembled Sequences

The general statistics for assembled plasmid sequences and hybrid assembled sequences are shown in Table 2. We observed a higher number and bigger contig size for *E. coli* than *K. pneumoniae* isolates. The GC percentage between *E. coli* and K. pneumoniae was similar, and the N50 values (i.e., the minimum contig length required to cover 50% of the assembled genome sequence) were higher in *E. coli*, indicating larger contig size of plasmids which denotes good quality of assembly. Regarding hybrid assembled isolates, generally, bigger contigs and higher N50 values were observed (Table 2). Interestingly GC percentage was higher in both *K. pneumoniae* isolates compared to the *E. coli* isolate.

Table 2. An overview of general statistics (mean ± SD) obtained using the QUAST tool for the Scheme 39. and *K. pneumoniae* 23, 37).

	Number of Contigs	Largest Contig (bp)	Total Length (bp)	GC (%)	N50
E. coli	87.5 ± 484.8	47,992 ± 65,592	271,880 ± 634,155	48.8 ± 2.3	31,029 ± 62,333
K. pneumoniae	7.1 ± 7.9	35,135 ± 48,184	98,095 ± 170,354	48.5 ± 3.8	23,022 ± 29,064
E. coli **39**	60	2,757,734	5,955,163	50.51	996,338
K. pneumoniae **23**	15	5,305,106	5,831,976	56.75	5,305,106
K. pneumoniae **37**	28	2,952,449	5,558,213	57.59	2,952,449

3.2. In Silico Plasmid Validation

Assembled plasmid sequences were further validated with the PlasmidFinder online tool and using BLASTn against PLSDB and Brooks plasmid database. Using the PlasmidFinder tool, we identified plasmid replicons in 39 (67%) and 11 (25%) of *E. coli* and *K. pneumoniae* isolates. This corresponds to two to three and one to two plasmid replicons per isolate caring plasmids in *E. coli* and *K. pneumoniae*, respectively. (Table S1).

The number of putative plasmids (P_{TRUE}) after BLASTn and removing duplicates hits per isolate for both *E. coli* and *K. pneumoniae* is shown in Figure 1A. Overall a higher number of P_{TRUE} was detected for *E. coli* than *K. pneumoniae* (Table S1). For *E. coli*, the majority of P_{TRUE} (122 of 173) detected in the Brooks database were detected using PLSDB as well. In *K. pneumoniae*, almost all the P_{TRUE} (29 of 30) detected using the Brooks database were also detected using the PLSDB database. Additionally, we detected 13 shared plasmids between *E. coli* and *K. pneumoniae* using the PLSDB database. In contrast, only three common plasmids were observed between *E. coli* and *K. pneumoniae* by employing the Brooks database. The Neighbor-Joining phylogenetic tree of IncFIB plasmids that were most prevalent for both *E. coli* and *K. pneumoniae* is presented in Figure 1B. The tree is based on MAFFT (Multiple Alignment using Fast Fourier Transform) alignment of plasmids conserved regions [28]. There were four IncFIB sequences of *K. pneumoniae*. Two *K. pneumoniae* isolates formed a separate branch, whereas two others were clustered together with IncFIB plasmid sequences from *E. coli*. An overview of the top 20 most abundant putative plasmids (P_{TRUE}) and replicons retrieved from each database for both *E. coli* and *K. pneumoniae* is shown in Table 3.

Regarding plasmid detection in hybrid assembled sequences, we managed to retrieve a higher number of P_{TRUE} from PLSDB than Brooks and PlasmidFinder. All plasmids detected in Brooks for *E. coli* 39 and *K. pneumoniae* 23 were also detected in the PLSDB database (Table S2). An overview of the top five P_{TRUE} and replicons retrieved from each database for hybrid assembled isolates can be seen in Table 4.
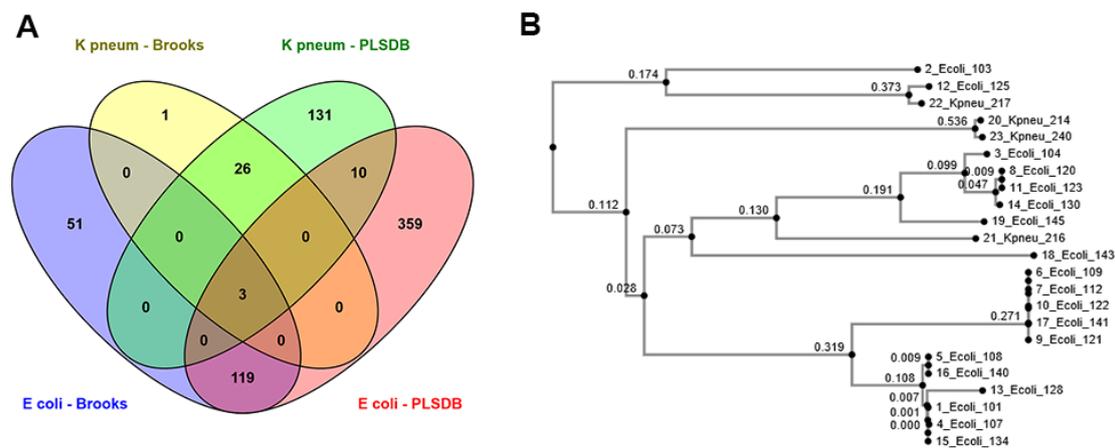
Figure 1. (**A**) Number of uniquely identified putative plasmids (P_{TRUE}) using different databases after removing duplicates hits per isolate, for both *E. coli* and *K. pneumoniae*. (**B**) phylogenic tree of IncFIB plasmids as the most prevalent plasmids for both *E. coli* and *K. pneumoniae*.

Table 3. An overview of the top 20 most abundant putative plasmids (P_{TRUE}) in both *E. coli* and *K. pneumoniae* retrieved from different databases or tools. (*n*: number of isolates).

	PlasmidFinder		Brooks		PLSDB	
	Replicon Name	*n*	*Plasmid Name*	*n*	*Plasmid Name*	*n*
E. coli	IncFIB (AP001918)	23	pECO-bc6	5	pMTY17816_OXA48	7
	Col156	15	pUT189	4	pMTY17823_OXA48	7
	IncFIA	9	pUM146	4	p53_E-OXA48	7
	IncFII (29)	9	pRS218	4	pECO-bc6	5
	IncFII	8	pSF-166-1	4	unnamed 7	5
	Col (BS512)	4	pSaT040	4	pUTI89	4
	IncFII (pRSB107)	4	pZH063-1	4	pEC14_114	4
	Col (MG828)	3	pECOS88	3	pUM146	4
	Col8282	2	pMRY16-002_5	3	pRS218	4
	IncFIC (FII)	2	pAPEC-O78-ColV	3	pSF-166-1	4
	IncFII (pCoo)	2	pPSUO78_1	3	unnamed 1	4
	IncI1-I (Gamma)	2	pSF-088-1	3	pSaT040	4
	IncX4	2	pG749_1	3	pZH063_1	4
	Col (pHAD28)	1	pECAZ147_1	3	p0.1229_1	4
	FIA (pBK30683)	1	pECSF1	2	plasmid 2	4
	IncFIB (pB171)	1	pCE10A	2	unnamed	3
	IncM1	1	p3PCN033	2	pF2_18C_Col	3
	IncQ1	1	pMVAST0167_1	2	pDB4277	3
	IncX1	1	pKPN-7c3	2	pCERC4	3
	p0111	1	pEC732_6	2	pCERC5	3

Table 3. *Cont.*

	PlasmidFinder		Brooks		PLSDB	
	Replicon Name	*n*	*Plasmid Name*	*n*	*Plasmid Name*	*n*
K. pneumoniae	IncFIB (K)	4	pKpn23412-4	1	pEC-243	1
	IncFIA (HI1)	2	pK2044	1	pM206-OXA181	1
	IncFIB (pKPHS1)	2	pKCTC2242	1	pM513-OXA181	1
	IncHI1B (pNDM-MAR)	2	PCN061p4	1	pM517-OXA181	1
	repB	2	plasmid B	1	pM518-OXA181	1
	Col (KPHS6)	1	plasmid A	1	pTHC11-1	1
	Col440I	1	pRJF293	1	pTMTA63631	1
	Col440II	1	pRJF999	1	pCA08	1
	ColpVC	1	pNY9_3	1	pKpvK54	1
	IncFIB (AP001918)	1	pESBL931	1	pBH100 alpha	1
	IncFII	1	unnamed 1	1	pVir_095132	1
	IncFII (29)	1	unnamed	1	pSF07201	1
	IncFII (K)	1	unnamed 3	1	pYHCC	1
	IncX3	1	pKp_Goe_579-6	1	pKS22	1
			pKp_Goe_473-5	1	U25P002	1
			pKp_Goe_832-5	1	pFAM22321	1
			pKp_Goe_024-5	1	p183660	1
			pKp_Goe_026-5	1	pKP2442_7c331	1
			pKp_Goe_021-5	1	pOXA-181_29144	1
			pKp_Goe_304-5	1	pEC-NRS18	1

Table 4. An overview of the first top-five putative plasmids (P$_{TRUE}$) in hybrid assembled *E. coli* and *K. pneumoniae* isolates retrieved from the different databases.

|	Replicon Name from PlasmidFinder	Plasmid Name from Brooks	Plasmid Name from PLSDB
E. coli **39**	IncHI2 IncHI2A p0111	pCFSAN002069_01 p25155 PDM02 PDM04	pCFSAN002069_01 p25155 p2EC1-4 pSH1148_107 PDM04
K. pneumoniae **23**	IncFIB(pNDM-Mar) IncFIB(pQil) IncFII(K)	pKp_Goe_579-5 pKp_Goe_832-5 pKp_Goe_304-5 pKp_Goe_021-5 pKp_Goe_026-5	p5 pEC25-4 p4_VBA2172 pRGI01215 p1_040074
K. pneumoniae **37**	IncFII IncFII(pKP91) IncFIB(K) IncFIA(HI1)		U25P002

3.3. Identification of Plasmid-Mediated ARGs

Plasmid assembled files were used to explore the plasmid-mediated ARGs using ResFinder and CARD databases. As can be seen from Table 5 and using plasmid data, regardless of whether identified ARGs located on the putative plasmids (P_{TRUE}) or not, we identified more ARGs using the CARD database in *E. coli* isolates compared to ResFinder. For *K. pneumoniae,* opposite results were observed. Moreover, several predicted ARGs were different after annotating the results to plasmid databases (for *E. coli* isolates, $ARG_{Plasmid-PlasmidFinder} > ARG_{Plasmid-PLSDB} > ARG_{Plasmid-Brooks}$ and for *K. pneumoniae* isolates, $ARG_{Plasmid-PLSDB} > ARG_{Plasmid-PlasmidFinder} > ARG_{Plasmid-Brooks}$). Moreover, using hybrid assembled data for *E. coli* 39 isolates, we detected a higher number of ARGs using the CARD database. The same results for *K. pneumoniae* 37 were also observed (Table S2). Surprisingly, no ARGs were detected using *K. pneumoniae* plasmid or hybrid data as $ARG_{Plasmid-Brooks}$.

Table 5. Number of total identified ARGs and $ARG_{Plasmid}$ (sharing the same contig with P_{TRUE} from three different databases), using ResFinder and CARD databases in both *E. coli* and *K. pneumoniae* and hybrid assembled isolates. The number inside parentheses shows the percentage of $ARG_{Plasmid}$ identified in P_{TRUE} from different plasmid databases. *E. coli* 39, *K. pneumoniae* 23, and *K. pneumoniae* 37 are hybrid assembled isolates.

	E. coli		*K. pneumoniae*		*E. coli* **39**		*K. pneumoniae* **23**		*K. pneumoniae* **37**	
	Res Finder	**CARD**	**Res Finder**	**CARD**	**Res Finder**	**CARD**	**Res Finder**	**CARD**	**Res Finder**	**CARD**
Total number of ARGs	61	100	29	15	9	54	24	24	9	17
PlasmidFinder	31 (51%)	27 (27%)	1 (3%)	1 (6%)	6 (54%)	6 (11%)	14 (58%)	12 (50%)	6 (54%)	5 (29%)
Brooks	10 (16%)	10 (10%)	-	-	2 (22%)	1 (2%)	-	-	-	-
PLSDB	18 (29%)	16 (16%)	3 (10%)	2 (13%)	7 (77%)	7 (13%)	7 (29%)	6 (25%)	6 (54%)	5 (29%)

As is apparent from Table 5, $ARG_{Plasmid-PlasmidFinder}$ in *E. coli* plasmid data represent a group with the highest detected number of ARGs. Further details about $ARG_{Plasmid-PlasmidFinder}$ can be seen in Figure 2 and Table 5. For *E. coli* plasmid data, the majority of $ARG_{Plasmid}$ were found on the IncFII plasmid. Furthermore, plasmids such as Col(pHAD28) and IncI1-1(Gamma) hosted the least ARGs. Some ARGs such as *aph(3″)-Ib*, *aph(6)-Id*, *blaTEM-1B*, and *sul2* were carried by more than one type of plasmid.
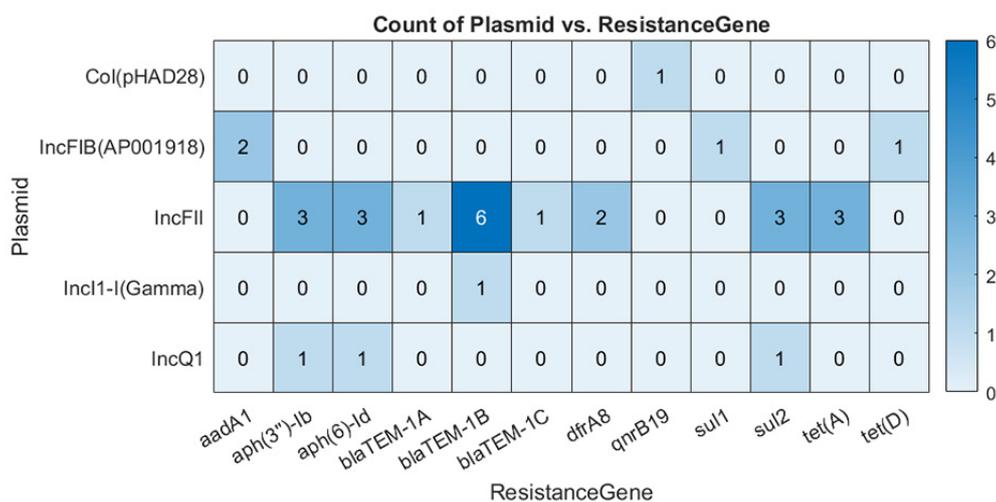
Figure 2. The co-existence of $ARG_{Plasmid}$ genes and different plasmids detected by PlasmidFinder in plasmid data from *E. coli* isolates. Numbers inside each cell indicating the number of isolates where ARGs were found on the corresponding plasmid.

The most abundant $ARG_{Plasmid}$ genes for plasmid data from *E. coli* and *K. pneumoniae* isolates can be seen in Table 6. The majority of detected $ARG_{Plasmid}$ hits in *E. coli* isolates, carried by P_{TRUE} from PlasmidFinder and PLSDB, were beta-lactamase gene-variants blaTEM-1B and TEM-1B. For *E. coli*, $ARG_{Plasmid}$ genes such as *aph(3″)-Ib*, *aph(6)-Id*, *sul1*, *sul2*, and *tet(D)* were flagged as mutual $ARG_{Plasmid}$, observed in all databases. For *K. pneumoniae*, no $ARG_{Plasmid}$ gene was detected on P_{TRUE} from Brooks. However, the *qnrS1* gene was found as a mutual $ARG_{Plasmid}$ harbored by P_{TRUE} from both PLSDB and PlasmidFinder.

The $ARG_{plasmid}$ prediction using hybrid assembled sequences is presented in Table 7. Overall, $ARG_{plasmid}$ prediction using hybrid assembled sequences was more consistent between databases compared to plasmid assembled data (Table S2). For instance, in the hybrid assembled *E. coli* 39 isolate, $ARG_{Plasmid}$ such as *aac(3)-VIa* and *aadA1* were hosted by P_{TRUE} from all databases. For *K. pneumoniae* isolate 37, predicted $ARG_{Plasmid-PlasmidFinder}$ and $ARG_{Plasmid-PLSDB}$ were entirely matched. Regarding *K. pneumoniae* isolate 23, besides an extra predicted $ARG_{Plasmid-PlasmidFinder}$, all the predicted $ARG_{Plasmid-PLSDB}$ were covered by $ARG_{Plasmid-PlasmidFinder}$.

Table 6. Gene name and number of isolates with most abundant $ARG_{Plasmid}$ for both *E. coli* and *K. pneumoniae* isolates.

| | *E. coli* | | | | *K. pneumoniae* | | | |
| | **ResFinder** | | **CARD** | | **ResFinder** | | **CARD** | |
	Gene Name	*n*	**Gene Name**	*n*	**Gene Name**	*n*	**Gene Name**	*n*
PlasmidFinder	aadA1	2	aadA	1	qnrS1	1	qnrS1	1
	aph(3″)-Ib	4	aph(3″)-Ib	3				
	aph(6)-Id	4	aph(6)-Id	4				
	blaTEM-1A	1	TEM-1	8				
	blaTEM-1B	7	TEM-40	1				
	blaTEM-1C	1						
	dfrA8	2	dfrA8	2				
	qnrB19	1	qnrB19	1				
	sul1	1	sul1	1				
	sul2	4	sul2	4				
	tet(A)	3	qacEdelta1	1				
	tet(D)	1	tet(D)	1				
Brooks	aadA1	1	aadA	1				
	aph(3″)-Ib	2	aph(3″)-Ib	1				
	aph(6)-Id	2	aph(6)-Id	2				
	blaTEM-1B	1	TEM-1	1				
	sul1	1	sul1	1				
	sul2	2	sul2	2				
	tet(D)	1	tet(D)	1				
			qacEdelta1	1				
PLSDB	aadA1	1	aadA	1	qnrS1	1	qnrS1	1
	aadA5	1	aadA5	1	floR	1	SHV-1	1
	aph(3″)-Ib	2	aph(3″)-Ib	1	blaSHV-99	1		
	aph(6)-Id	2	aph(6)-Id	2				
	blaTEM-1B	3	TEM-1	3				
	blaTEM-1D	1						
	dfrA17	1						
	mph(A)	1	mph(A)	1				
	sul1	2	sul1	2				
	sul2	2	sul2	2				
	tet(A)	1	qacEdelta1	2				
	tet(D)	1	tet(D)	1				

Table 7. Predicted ARG$_{plasmid}$ using hybrid assembled data for both *E. coli* and *K. pneumoniae* isolates.

	E. coli 39		*K. pneumoniae* 23		*K. pneumoniae* 37	
	ResFinder	**CARD**	**ResFinder**	**CARD**	**ResFinder**	**CARD**
PlasmidFinder	aac(3)-VIa aadA1	aadA	aac(3)-IId aac(6′)-Ib aac(6′)-Ib-cr aadA1 aph(3′)-Ia aph(3″)-Ib aph(6)-Id blaCTX-M-15 blaOXA-9 blaSHV-12 blaTEM-1A blaTEM-1B catA1 dfrA30 sul2	aac(3)-IId aac(6′)-Ib10 aadA aph(3′)-Ia aph(6)-Id CTX-M-15 OXA-9 SHV-134 TEM-1 catI qacE sul2	aac(3)-IIa aac(6′)-Ib3 aac(6′)-Ib-cr blaCTX-M-14 blaTEM-1B cmlA1	aac(3)-IIe CTX-M-14 TEM-1 cmlA5 qacEdelta1
Brooks	aac(3)-VIa aadA1					
PLSDB	aac(3)-VIa aadA1 blaCTX-M-2 blaTEM-1B dfrA1 sul1 tet(A)	aadA CTX-M-2 TEM-1 dfrA1 sul1 qacEdelta1	aac(6′)-Ib aac(6′)-Ib-cr aadA1 blaCTX-M-15 blaOXA-9 blaSHV-12 blaTEM-1A	aac(6′)-Ib10 aadA CTX-M-15 OXA-9 SHV-134 TEM-1	aac(3)-IIa aac(6′)-Ib3 aac(6′)-Ib-cr blaCTX-M-14 blaTEM-1B cmlA1	aac(3)-IIe CTX-M-14 TEM-1 cmlA5 qacEdelta1

4. Discussion

In the current research, the applicability of three different plasmid databases and two antibiotics resistance gene databases were assessed using *E. coli* and *K. pneumoniae* assemblies taken from Norwegian patients.

We identified a total number of 490 and 52 exclusive putative plasmids using PLSDB and Brooks databases, respectively. Observed differences might be explained by the content of databases, as the method used for developing the databases and the date of last revision (October 2018 for Brooks and November 2020 for PLSDB) as well as their file size (11,677 and 13,789 entries in Brooks and PLSDB, respectively) are different. Although a BLASTn search against Brooks and PLSDB databases resulted in a higher number of putative plasmids than PlasmidFinder, the method has its disadvantages. For instance, using the BLASTn search, we have detected multiple hits with similar lengths, alignment coverage, and percentage identity for the same assigned contig. Therefore, assigning the putative plasmids as P_{TRUE} was challenging. Similar challenges following BLAST+ have been previously described for the FindPlasmid package [29]. Using PlasmidFinder, researchers can directly upload raw files from sequencing platforms. Therefore, de novo assembly is not required, and PlasmidFinder can perform de novo assembly automatically, though the assembly results are not presented by the tool. On the other hand, manual de novo assembly is required in advance to BLASTn search when using other databases such as Brooks and PLSDB. However, one of the disadvantages of using PlasmidFinder is that it currently only covers *Enterobacteriaceae* and a few Gram-positive bacterial species.

It is clinically relevant to perform downstream analyses such as the prediction of plasmid associated ARGs following plasmid identification. In this study, ResFinder performed better than CARD to predict plasmid associated antibacterial resistance genes (AMR$_{Plasmid-PlasmidFinder}$, AMR$_{Plasmid-Brooks}$, and AMR$_{Plasmid-PLSDB}$) for both plasmid and hybrid assembled data. In a study comparing the performance of resistance gene databases, both CARD and ResFinder performed equally when submitting a single gene sequence,

but CARD performed slightly better for assembled data [30]. Although CARD only accepts FASTA assembly files up to 20 Mb, but in addition to acquired gene information, it contains chromosomal mutation data too. However, ResFinder takes raw files, and assembly is not required. Furthermore, in ResFinder, users can choose between acquired genes or chromosomal mutations. One of the ResFinder advantages is flagging the hit with the true circular term, which indicates whether the hit is plasmid associated or not. Therefore, current data suggest using PlasmidFinder and its associated ResFinder online tools as the first choice to predict plasmid associated ARGs.

In the current study, the AMR$_{Plasmid}$ gene profiles differed between *E. coli* strains carrying plasmids of the same type. Similar results have been reported for *Salmonella entrica* isolates in Ghana [31]. This further highlights the mobility of genetic elements between plasmids, resulting in acquiring or losing the ability for antimicrobial resistance. IncF plasmids are known carriers of a broad spectrum of antibiotic resistance genes in *E. coli* [32–34]. In line with this, IncFII plasmids were strongly associated with various resistance genes in our study. These plasmids carried TEM-1B, *aph*, *sul*, *tetA*, and *dfr* genes conferring resistance to penicillins, aminoglycosides, sulfonamides, tetracyclines, and trimethoprim [35]. IncFIB were the most prevalent plasmids in our dataset, and they exhibited a low association with antibiotic resistance genes. As such, *aad* (aminoglycoside), *sul* (sulfonamides), and *tet* (tetracycline) were located on IncFIB plasmid contig in two cases, highlighting the low, albeit growing, antibiotic resistance in Norway [16]. However, phylogenetic analysis of IncFIB plasmids revealed that plasmid sequences were shared between *E. coli* and *K. pneumoniae*, probably indicating its ability for inter-species transfer, which raises a concern over rising antibiotic resistance in Norway [17]. Additionally, in this work, we documented the co-existence of blaCTX−M genes with other genes corresponding to resistance against sulfonamide, aminoglycoside, trimethoprim, and tetracycline. This agrees with previous reports indicating that plasmids harboring blaCTX−M genes frequently also carry other genes encoding resistance to other antimicrobials [36–38].

The high sequence error rate in Oxford Nanopore Technologies and incongruity between short/fragmented reads from MiSeq Illumina platform and large repetitive regions in plasmids often results in the inaccurate prediction of plasmid-mediated ARGs. To overcome this issue, hybrid assembly has been suggested [12,39]. In the present research, we found that the prediction of ARG$_{Plasmids}$ following hybrid assembly was more consistent across different databases. Having a less fragmented assembly where the circular plasmids are apparent makes the prediction of ARG$_{plasmid}$ more accurate. The current conclusion regarding the applicability of hybrid assembly for plasmid-mediated ARGs detection previously has been made [40]. Therefore, future work implanting hybrid assembly to identify ARGs in bacteria is worth investigating.

5. Conclusions

In conclusion, we have demonstrated that plasmid detection and plasmid-mediated ARG prediction are challenging and to obtain a reliable result, one must consider different tools and databases. In the present study, a combination of PlasmidFinder and ResFinder tools showed promising results for both *E. coli* and *K. pneumoniae* isolates. Plasmid detection and prediction of plasmid-mediated ARG can be facilitated using hybrid assembly. Although Norway is considered as a country with a low antibiotic resistance frequency, current research provides a reasonable argument to tackle the slightly increasing antibiotic resistance issue in Norway.

References

1. Cassini, A.; Högberg, L.D.; Plachouras, D.; Quattrocchi, A.; Hoxha, A.; Simonsen, G.S.; Colomb-Cotinat, M.; E Kretzschmar, M.; Devleesschauwer, B.; Cecchini, M.; et al. Attributable deaths and disability-adjusted life-years caused by infections with antibiotic-resistant bacteria in the EU and the European Economic Area in 2015: A population-level modelling analysis. *Lancet Infect. Dis.* **2019**, *19*, 56–66. [CrossRef]
2. WHO. *Global Priority List of Antibiotic-Resistant Bacteria to Guide Research, Discovery, and Development of New Antibiotics*; WHO: Geneva, Switzerland, 2017.
3. Dunn, S.J.; Connor, C.; McNally, A. The evolution and transmission of multi-drug resistant Escherichia coli and Klebsiella pneumoniae: The complexity of clones and plasmids. *Curr. Opin. Microbiol.* **2019**, *51*, 51–56. [CrossRef] [PubMed]
4. Von Wintersdorff, C.J.; Penders, J.; Van Niekerk, J.M.; Mills, N.D.; Majumder, S.; Van Alphen, L.B.; Savelkoul, P.H.; Wolffs, P.F. Dissemination of Antimicrobial Resistance in Microbial Ecosystems through Horizontal Gene Transfer. *Front. Microbiol.* **2016**, *7*, 173. [CrossRef]
5. Bennett, P.M. Plasmid encoded antibiotic resistance: Acquisition and transfer of antibiotic resistance genes in bacteria. *Br. J. Pharmacol.* **2008**, *153*, S347–S357. [CrossRef]
6. Kaper, J.B.; Nataro, J.P.; Mobley, H.L.T. Pathogenic Escherichia coli. *Nat. Rev. Microbiol.* **2004**, *2*, 123–140. [CrossRef]
7. Hiraga, S.; Sugiyama, T.; Itoh, T. Comparative analysis of the replicon regions of eleven ColE2-related plasmids. *J. Bacteriol.* **1994**, *176*, 7233–7243. [CrossRef]
8. Köser, C.U.; Ellington, M.J.; Cartwright, E.J.P.; Gillespie, S.H.; Brown, N.M.; Farrington, M.; Holden, M.T.G.; Dougan, G.; Bentley, S.D.; Parkhill, J.; et al. Routine use of microbial whole genome sequencing in diagnostic and public health microbiology. *PLoS Pathog.* **2012**, *8*, e1002824. [CrossRef]
9. Arredondo-Alonso, S.; Rogers, M.R.C.; Braat, J.C.; Verschuuren, T.D.; Top, J.; Corander, J.; Willems, R.J.L.; Schürch, A.C. Mlplasmids: A user-friendly tool to predict plasmid- and chromosome-derived sequences for single species. *Microbial. Genom.* **2018**, *4*, e000224. [CrossRef]
10. Tolmasky, M.E.; Alonso, J.C. *Plasmids: Biology and Impact in Biotechnology and Discovery*; ASM Press: Washington, DC, USA, 2015.
11. Lal Gupta, C.; Kumar Tiwari, R.; Cytryn, E. Platforms for elucidating antibiotic resistance in single genomes and complex metagenomes. *Environ. Int.* **2020**, *138*, 105667. [CrossRef]
12. Taxt, A.M.; Avershina, E.; Frye, S.A.; Naseer, U.; Ahmad, R. Rapid identification of pathogens, antibiotic resistance genes and plasmids in blood cultures by nanopore sequencing. *Sci. Rep.* **2020**, *10*, 7622. [CrossRef]
13. Harstad, H.; Ahmad, R.; Bredberg, A. Nanopore-based DNA sequencing in clinical microbiology: Preliminary assessment of basic requirements. *arXiv* **2018**, arXiv:bio/382580.
14. Yang, Y.; Stathis, D.; Sharma, P.; Paul, K.; Hemani, A.; Grabherr, M.; Ahmad, R. RiBoSOM: Rapid bacterial genome identification using self-organizing map implemented on the synchoros SiLago platform. In *ACM International Conference Proceeding Series*; Association for Computing Machinery (ACM): Pythagorion, Greece, 2018; pp. 105–114.

15. Stathis, D.; Yang, Y.; Tewari, S.; Hemani, A.; Paul, K.; Grabherr, M.; Ahmad, R. Approximate Computing Applied to Bacterial Genome Identification using Self-Organizing Maps. In Proceedings of the 2019 IEEE Computer Society Annual Symposium on VLSI (ISVLSI), Miami, FL, USA, 15–17 July 2019; pp. 560–567. [CrossRef]

16. Klein, E.Y.; Tseng, K.K.; Pant, S.; Laxminarayan, R. Tracking global trends in the effectiveness of antibiotic therapy using the Drug Resistance Index. *BMJ Glob. Health* **2019**, *4*, e001315. [CrossRef] [PubMed]

17. NORM/NORM-VET. *Usage of Antimicrobial Agents and Occurrence of Antimicrobial Resistance in Norway*; Norwegian Veterinary Institute: Tromsø/Oslo, Norway, 2019.

18. Avershina, E. AMR-Diag: Neural Network based Genotype-to-Phenotype Prediction of Resistance towards Beta-Lactams and Colistin in Escherichia Coli and Klebsiella Pneumoniae. *Comput. Struct. Biotechnol. J. Elsevier* **2020**. under review.

19. Andrews, S. FastQC: A Quality Control. Tool for High. Throughput Sequence Data. Available online: http://www.bioinformatics.babraham.ac.uk/projects/fastqc/ (accessed on 24 September 2020).

20. Bolger, A.M.; Lohse, M.; Usadel, B. Trimmomatic: A flexible trimmer for Illumina sequence data. *Bioinformatics* **2014**, *30*, 2114–2120. [CrossRef]

21. Antipov, D.; Hartwick, N.; Shen, M.; Raiko, M.; Lapidus, A.; Pevzner, P.A. PlasmidSPAdes: Assembling plasmids from whole genome sequencing data. *Bioinformatics* **2016**, *32*, 3380–3387. [CrossRef]

22. Gurevich, A.; Saveliev, V.; Vyahhi, N.; Tesler, G. QUAST: Quality assessment tool for genome assemblies. *Bioinformatics* **2013**, *29*, 1072–1075. [CrossRef]

23. Carattoli, A.; Zankari, E.; García-Fernández, A.; Larsen, M.V.; Lund, O.; Villa, L.; Aarestrup, F.M.; Hasman, H. In silico detection and typing of plasmids using PlasmidFinder and plasmid multilocus sequence typing. *Antimicrob. Agents Chemother.* **2014**, *58*, 3895–3903. [CrossRef]

24. Galata, V.; Fehlmann, T.; Backes, C.; Keller, A. PLSDB: A resource of complete bacterial plasmids. *Nucleic Acids Res.* **2019**, *47*, D195–D202. [CrossRef]

25. Brooks, L.; MoSistrom, M. A Curated, Comprehensive Database of Plasmid Sequences. *Microbiol. Resour. Announc.* **2019**, *8*, e0132518. [CrossRef]

26. Bortolaia, V.; Ruppe, R.S.; Roberts, E.; Schwarz, M.C.; Cattoir, S.; Philippon, V.; Allesoe, A.; Rebelo, R.L.; Florensa, A.R.; Fagelhauer, A.F.; et al. ResFinder 4.0 for predictions of phenotypes from genotypes. *J. Antimicrob. Chemother.* **2020**, *75*, 3491–3500. [CrossRef]

27. Alcock, B.P.R.; Lau, A.R.; Tsang, T.T.Y.; Bouchard, K.K.; Edalatmand, M.; Huynh, A.; Nguyen, W.; Cheng, A.V.; Liu, A.A.; Min, S.; et al. CARD 2020: Antibiotic resistome surveillance with the comprehensive antibiotic resistance database. *Nucleic Acids Res.* **2020**, *48*, D517–D525. [CrossRef] [PubMed]

28. Katoh, K.; Rozewicki, J.; Yamada, K.D. MAFFT online service: Multiple sequence alignment, interactive sequence choice and visualization. *Brief. Bioinform.* **2017**, *20*, 1160–1166. [CrossRef] [PubMed]

29. Hall, B.G. Identifying Plasmids in Bacterial Genome Assemblies. *arXiv* **2018**, arXiv:bio/332049.

30. Xavier, B.B.; Das, A.J.; Cochrane, G.; De Ganck, S.; Kumar-Singh, S.; Aarestrup, F.M.; Goossens, H.; Malhotra-Kumar, S. Consolidating and Exploring Antibiotic Resistance Gene Data Resources. *J. Clin. Microbiol.* **2016**, *54*, 851–859. [CrossRef]

31. Kudirkiene, E.; Andoh, L.A.; Ahmed, S.; Herrero-Fresno, A.; Dalsgaard, A.; Obiri-Danso, K.; Olsen, J.E. The Use of a Combined Bioinformatics Approach to Locate Antibiotic Resistance Genes on Plasmids from Whole Genome Sequences of Salmonella enterica Serovars from Humans in Ghana. *Front. Microbiol.* **2018**, *9*, 1010. [CrossRef]

32. Lyimo, B.; Buza, J.; Subbiah, M.; Temba, S.; Kipasika, H.; Smith, W.; Call, D.R. IncF Plasmids Are Commonly Carried by Antibiotic Resistant *Escherichia coli* Isolated from Drinking Water Sources in Northern Tanzania. *Int. J. Microbiol.* **2016**, *2016*, 3103672. [CrossRef]

33. Torres-González, P.; Bobadilla-del Valle, M.; Tovar-Calderon, E.; Leal-Vega, F.; Hernandez-Cruz, A.; Martinez-Gamboa, A.; Niembro-Ortega, M.D.; Martinez-Gamboa, A.; Ponce-de-Leon, A. Outbreak Caused by Enterobacteriaceae Harboring NDM-1 Metallo-β-Lactamase Carried in an IncFII Plasmid in a Tertiary Care Hospital in Mexico City. *Antimicrob. Agents Chemother.* **2015**, *59*, 7080–7083. [CrossRef]

34. Deng, Y.; Zeng, Z.; Chen, S.; He, L.; Liu, Y.; Wu, C.; Chen, Z.; Yao, Q.; Hou, J.; Yang, T.; et al. Dissemination of IncFII plasmids carrying rmtB and qepA in Escherichia coli from pigs, farm workers and the environment. *Clin. Microbiol. Infect.* **2011**, *17*, 1740–1745. [CrossRef]

35. Van Hoek, A.H.; Mevius, D.; Guerra, B.; Mullany, P.; Roberts, A.P.; Aarts, H.J. Acquired Antibiotic Resistance Genes: An. *Overview. Front. Microbiol.* **2011**, *2*, 203.

36. Liu, B.T.; Yang, Q.E.; Li, L.; Sun, J.; Liao, X.P.; Fang, L.X.; Yang, S.S.; Deng, H.; Liu, Y.H. Dissemination and characterization of plasmids carrying oqxAB-bla CTX-M genes in Escherichia coli isolates from food-producing animals. *PLoS ONE* **2013**, *8*, e73947.

37. Sun, H.; Li, S.; Xie, Z.; Yang, F.; Zhu, Y.; Zhao, X.; Jiang, S. A novel multidrug resistance plasmid isolated from an Escherichia coli strain resistant to aminoglycosides. *J. Antimicrob. Chemother.* **2012**, *67*, 1635–1638. [CrossRef] [PubMed]

38. Huang, W.; Wang, G.; Sebra, R.; Zhuge, J.; Yin, C.; Aguero-Rosenfeld, M.E.; Schuetz, A.N.; Dimitrova, N.; Fallon, J.T. Emergence and Evolution of Multidrug-Resistant Klebsiella pneumoniae with both bla(KPC) and bla(CTX-M) Integrated in the Chromosome. *Antimicrob. Agents Chemother.* **2017**, *61*, e00076-17. [CrossRef] [PubMed]

39. Bertrand, D.; Shaw, J.; Kalathiyappan, M.; Ng, A.H.Q.; Kumar, M.S.; Li, C.; Dvornicic, M.; Soldo, J.P.; Koh, J.Y.; Tong, C.; et al. Hybrid metagenomic assembly enables high-resolution analysis of resistance determinants and mobile elements in human microbiomes. *Nat. Biotechnol.* **2019**, *37*, 937–944. [CrossRef] [PubMed]
40. Berbers, B.; Ceyssens, P.-J.; Bogaerts, P.; Vanneste, K.; Roosens, N.H.; Marchal, K.; De Keersmaecker, S.C.J. Development of an NGS-Based Workflow for Improved Monitoring of Circulating Plasmids in Support of Risk Assessment of Antimicrobial Resistance Gene Dissemination. *Antibiotics* **2020**, *9*, 503. [CrossRef] [PubMed]